REPORT RESUMES

ED 003 886 48
EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION--VI. PROGRESS REPORT NO. 6.
BY- LANE, HARLAN
MICHIGAN UNIV., ANN ARBOR,COLL.OF LIT.,SCI.,ARTS
REPORT NUMBER NDEA-VI-78-6 PUB DATE NOV 64
REPORT NUMBER PR-6
REPORT NUMBER 05586-05613-3-P
CONTRACT OEC-3-14-013
EDRS PRICE MF-$0.45 HC-$12.88 322P.

DESCRIPTORS- *SPEECH, *PERCEPTUAL DEVELOPMENT, *LANGUAGE RESEARCH, *EXPERIMENTAL PROGRAMS, THEORIES, *AUTOINSTRUCTIONAL AIDS, ANN ARBOR, MICHIGAN

VARIOUS ASPECTS OF THE PROGRESS OF AN EXPERIMENTAL PROGRAM IN SPEECH CONTROL WERE REPORTED. THE TOPICS COVERED IN THE DISCUSSIONS WERE (1) ACOUSTIC AND DISCRIMINATIVE PROPERTIES OF SPEECH SOUNDS, (2) MATCHING FUNCTIONS AND EQUAL-SENSATION CONTOURS FOR LOUDNESS, (3) RELATIONS BETWEEN IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR SPEECH AND NONSPEECH CONTINUUMS, (4) IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR GRAPHEMES AND LINES, (5) A RELATION BETWEEN STIMULUS GENERALIZATION AND MAGNITUDE ESTIMATION, (6) AN APPLICATION OF MEAN VALUE THEORY TO PSYCHOPHYSICAL MEASUREMENT, (7) SELF-INSTRUCTIONAL DEVICE FOR CONDITIONING ACCURATE PROSODY, AND (8) PROGRAM FOR REVIEWS, AND A REVIEW OF A PROGRAM ON LINGUISTICS. (RELATED REPORTS ARE ED 003 883-ED 003 887.) (RS)

ED003886

78
No 6

05586, 05613-3-P

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
DEPARTMENT OF PSYCHOLOGY
BEHAVIOR ANALYSIS LABORATORY

Progress Report No. 6

Experimental Analysis of the Control of Speech Production and Perception: VI

Supported by:

Language Development Section
U. S. Office of Education
Contract No. OE-3-14-013
Washington, D. C.

and

National Science Foundation
Grant No. GB-552
Washington, D. C.

Administered through: *October 1964*

OFFICE OF RESEARCH ADMINISTRATION · ANN ARBOR

THE UNIVERSITY OF MICHIGAN
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Department of Psychology
Behavior Analysis Laboratory

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Progress Report No. 6

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION: VI

ORA Projects 05613 and 05586

supported by:

LANGUAGE DEVELOPMENT SECTION
U. S. OFFICE OF EDUCATION
CONTRACT NO. OE-3-14-013
WASHINGTON, D. C.

and

NATIONAL SCIENCE FOUNDATION
GRANT NO. GB-552
WASHINGTON, D.C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

November 1964

PRECEDING PAGE BLANK-NOT FILMED

TABLE OF CONTENTS

STAFF

A SURVEY OF THE ACOUSTIC AND DISCRIMINATIVE PROPERTIES OF SPEECH SOUNDS

MATCHING FUNCTIONS AND EQUAL-SENSATION CONTOURS FOR LOUDNESS

AN ANALYSIS OF THE RELATIONS BETWEEN IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR SPEECH AND NONSPEECH CONTINUA

IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR GRAPHEMES AND LINES

A RELATION BETWEEN STIMULUS GENERALIZATION AND MAGNITUDE ESTIMATION

AN APPLICATION OF MEAN VALUE THEORY TO PSYCHOPHYSICAL MEASUREMENT

A SELF-INSTRUCTIONAL DEVICE FOR CONDITIONING ACCURATE PROSODY

A PROGRAM FOR REVIEWS AND A REVIEW OF A PROGRAM ON LINGUISTICS

PRECEDING PAGE BLANK-NOT FILMED

STAFF

# of the

BEHAVIOR ANALYSIS LABORATORY

H. L. Lane, Ph.D., Director

S. A. Ross, Mag. Art., Research Associate

D. V. Cross, A.M., Research Assistant

J. L. Kopp, A.M., Research Assistant

R. L. Buiten, B.S.E., Research Engineer

D. R. Brinkman, B.S.E., Administrative Assistant

W. C. Sheppard, B.A., Research Assistant

R. E. Lehman, Technical Assistant

G. A. Lombard, Technical Assistant

D. M. Carlson, B.A., Programming Consultant

R. E. Skrocke, Research Technician

L. G. Peterson, Secretary

B. L. Keim, B.M., Typist

A SURVEY OF THE ACOUSTIC AND DISCRIMINATIVE PROPERTIES OF SPEECH SOUNDS

Harlan Lane
The University of Michigan

A Survey of the Acoustic and Discriminative Properties of Speech Sounds

Harlan Lane

Behavior Analysis Laboratory

University of Michigan

The techniques and findings of descriptive linguistics and its allied discipline, phonetics, make it possible to describe the myriad and variegated utterances by the speakers of a language in terms of a relatively limited inventory of discrete speech sounds. The methods and instruments of acoustics and allied physical sciences make it possible to specify the physical properties of these speech sounds. The techniques of experimental psychology make it possible to identify the discriminative properties of these speech sounds--those physical differences that "make a difference" to the listener.

This survey describes the findings in a relatively new area of research, called acoustic phonetics, which draws on all three disciplines. An acquaintance with the basic terms and concepts of each discipline is assumed.[1] The survey is addressed to these two questions, with particular reference to American English: (1) What are the acoustic invariances that characterize the different sounds of speech? (2) Which of these acoustic invariances have a distinctive function; that is, which exert differential control over the behavior of listeners?

The vowels

Formant frequencies

Figure 1 presents a spectrogram of the utterance "Bell Telephone Laboratories" (the place where spectrographic techniques were pioneered by Potter, Kopp and Green, 1947, and Peterson, 1951a, among others). In contrast to the consonant segments in Fig. 1, the

----
Insert Fig. 1 about here
----

vowels are characterized by two or three energy maxima which have a relatively long duration and, typically, a brief "steady state", during which their frequency loci are constant. An examination of the acoustic properties of the vowels can best begin with isolated sustained vowels; we will return later to the effects of phonetic context, which are clearly marked in Fig. 1. Figure 2 presents spectrograms and amplitude sections of nine sustained vowels from midwest American English. The individual voice harmonics may be seen in the frequency x amplitude representations at the bottom of the Figure. The amplitude peaks in these spectral analyses reflect the reinforcement of voice harmonics by the resonances of the vocal tract and correspond to the formants in the frequency x time representations at the top of the Figure. Thus, the amplitude section of [i], measured at one point in time during the phonation of the vowel, shows one peak at low frequency and two at high frequency. The superposed spectrogram shows that this three-formant pattern is maintained, with approximately the same frequency values, throughout the duration of the sustained vowel.

- - - - - - - - - - - - - - - - - - -

Insert Fig. 2 about here

- - - - - - - - - - - - - - - - - - -

The vowels shown in Fig. 2 are represented by points in the acoustic vowel diagram of Fig. 3, according to the frequencies of their first and second formants. It is a common practice to place the origin of an acoustic vowel diagram at the upper right, so that the spatial arrangement of the vowel points corresponds roughly to their point of articulation in a vertical section of the oral cavity, as indicated by the directional labels. Thus [i], for example, which has a low first formant (F1) and a high second formant (F2) (cf. Fig. 2), and is articulated with the hump of the tongue relatively high and toward the front of the mouth, is represented by a point, with coordinates F2, F1, at a high and front (left) position in the acoustic vowel diagram. The frequency values on the coordinates are spaced linearly below and logarithmically above 1,000 cps, in accord with the Koenig scale of pitch (1949). The orderliness of the vowel diagram is a result of representing exclusively one sample of a few widely separated vowels, ut-

- - - - - - - - - - - - - - - - - - -

Insert Fig. 3 about here

- - - - - - - - - - - - - - - - - - -

tered in isolation by a single, trained speaker. We may now examine several sources of variance in the acoustic properties of vowels which confound the diagram but not, apparently, the listener. When the same vowel is uttered by a man, a woman and a child the three vowel phones will have very different formant frequencies, reflecting the differences in the size of the resonating cavities of the three speakers. This outcome is shown by the acoustic vowel diagram in Fig. 4. The [i] phones, for example, range over

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 4 about here
- - - - - - - - - - - - - - - - - - - -

1,000 cps in the frequencies of their second formant when they are uttered by the three different speakers. Another result of the physiological differences among the speakers is that different vowel phonemes have similar acoustic phones; the male [ɑ], for example, lies close to the female [ɔ] in the parametric representation of Fig. 4. Presumably, the listener effectively transforms these acoustic parameters during vowel perception. If we knew the transformation and applied it to the coordinates of Fig. 4, the phones of a single phoneme should cluster, and the clusters associated with different phonemes should be widely separated. This assumes, of course, that F1 and F2 are the only "information-bearing elements" (Peterson, 1952) of the vowels--an oversimplification which will be corrected later. The loops in Fig. 4 which connect the "outlying" vowels of each speaker indicate that there is a pattern invariance and that the vowel diagrams might be made congruent by normalizing the formant frequencies of the vowels for each speaker.[2] Inspection of Fig. 4 reveals that specifying the vowels in terms of the ratios of their formant frequencies will not yield the desired clustering, since the sets of three phones do not lie along separate straight lines in the graph. When a vowel diagram reflecting more sources of variance is considered, this solution is even less satisfactory (Peterson, 1961).

A likely candidate for reducing the variance illustrated in Fig. 4 is the frequency of the third formant, since F3 is also peculiar to the size of the vocal cavities of each speaker. Figure 5 shows that when the frequencies of the first two formants of a vowel

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 5 about here
- - - - - - - - - - - - - - - - - - - -

are normalized with respect to its third-formant frequency, the vowel phones uttered by man, woman and child cluster appropriately. (For this figure, the formant frequencies were transformed by the mel scale [Stevens & Volkman, 1940] and the origin was again placed at the upper right). Additional sources of variance continue to complicate the picture, however.

The variability in productions of a vowel on different days is illustrated in Fig. 6 for two speakers. Even when isolated, sustained vowels by one speaker are considered, a vowel phoneme is represented in an F1 x F2 plot as a region rather than as a point, and in some cases these regions overlap. The overlap due to this source of variance, illustrated by the phonemes [ɛ] and [æ] in the diagram at the left of Fig. 6, may be examined more closely with the measurements shown in Fig. 7 for a different speaker. The frequency distributions for F2 in repeated utterances of the two phonemes approach normality, indicating that the fluctuations reflect a host of minor and essentially random variables. The overlapping distributions for F1 present a similar picture.

- - - - - - - - - - - - - - - - - -
Insert Fig. 6 about here
- - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - -
Insert Fig. 7 about here
- - - - - - - - - - - - - - - - - -

When the two sources of variability which we have been considering, inter- and intra-speaker variance, are combined, and their effects are sampled recurrently, we find the more realistic picture of vowel acoustics shown in Fig. 8. A total of 76 speakers, men, women and children, spoke ten vowels (all in the context h___d) on two occasions. The loops in Fig. 8 enclose most of the phone points which were determined for the vowels, but only those points are shown here which correspond to vowels classified unanimously by a panel of 70 listeners. (For clarity, the phones spoken by women, which lie between those for the men and children, have not been plotted.) Figure 8 shows that the acoustic prop-

- - - - - - - - - - - - - - - - - -
Insert Fig. 8 about here
- - - - - - - - - - - - - - - - - -

erties of the different vowels overlap in their formant parameters, although listener identifications of the same vowels do not overlap. Table I presents the average parameter values measured for these vowels by Peterson and Barney (1952); these classic data have often served as a reference for other investigations and have largely been confirmed by subsequent research (see below).[3]

- - - - - - - - - - - - - - - - - -
Insert Table I about here
- - - - - - - - - - - - - - - - - -

If Peterson and Barney had included such utterances as "hide", "hoyed", and "hued" in their experiment, the complex acoustic vowel diagram would become even more complicated. As shown in Fig. 9, diphthong syllable nuclei such as [aI], [oI] and [ju] are characterized by changing formant frequencies throughout most of their extent, although a listener typically identifies each as a single vowel. It is clear from Fig. 9 that a diphthong would be represented in an acoustic vowel diagram by a directional track rather than a single point. Holbrook and Fairbanks (1962) had speakers recite the surnames Hay, High, Hoy, Hoe, Howe, and Hugh, and then measured the formant frequencies of each of the diphthongs at five points in time. Figure 10 shows the large and overlapping regions of the vowel space that encompassed all the measurements for each diphthong. The directional tracks through the space are based on median parameter values in each time sample for each diphthong.

- - - - - - - - - - - - - - - - - -
Insert Fig. 9 about here
- - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - -
Insert Fig. 10 about here
- - - - - - - - - - - - - - - - - -

The phonetic environment of a vowel provides a further source of variance in its acoustic properties. The marked effects of adjacent consonants on the acoustic properties of an intervening vowel nucleus can readily be observed in the sample of speech represented by Fig. 1. In the case of some segments, the consonant-vowel transitions override the vowel steady-state, although "target" values for the formants may be identified.

Figure 11a shows the formant transitions of a vowel nucleus, measured by the technique of spectrum matching, and Fig. 11b shows how a target frequency may be identified in terms of a fit of a parabolic segment to the transitions. When the target frequencies of the first and second formants of vowels were measured in a wide variety of consonantal contexts, the average values shown in Table II were obtained. In terms of the acoustic vowel diagram, the effect of varying the consonantal context of a vowel is to scatter the points representing the vowel phones. This is shown in Fig. 12, where each point corresponds to the parameter values of a vowel, averaged over repeated utterances, in one of 14 consonantal environments, by one of three speakers. For comparison with the vowel space obtained when the consonantal environment is held constant, the vowel regions shown in Fig. 8 are also indicated in Fig. 12. The overall downward shift in the frequency of the second formant effected by the consonant environments is particularly conspicuous for the front vowels represented in the diagram (compare Tables I and II). The changes in the frequency of the first formant produced by varying the consonant environment of a vowel are comparatively quite small.

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 11 about here

- - - - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - - - -

Insert Table II about here

- - - - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 12 about here

- - - - - - - - - - - - - - - - - - - -

Instead of measuring the target positions of the formants of vowels in consonantal environments, which are shown in Table II and Fig. 12, Shearme and Holmes (1962) represented the vowels more completely as directional tracks through the acoustic vowel space--as was done with the diphthongs in Fig. 10. A few tracks for the vowel [i] are shown in Fig. 13 along with that vowel's parameter values when uttered in an isolated monosyllable.

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 13 about here

- - - - - - - - - - - - - - - - - - - -

In order to associate the vowel tracks with a minimal region of the acoustic space, the smallest area which contains part of each track is described by the dotted line. When a set of these vowel regions is prepared for the utterances of one speaker, the vowel space shown in Fig. 14 is obtained. If the regions for other speakers were introduced and broadened somewhat to encompass more of the vowel tracks observed in the stream of speech, the acoustic vowel space would be virtually filled by a rat's nest of overlapping phoneme regions.

- - - - - - - - - - - - - - - - - - -

Insert Fig. 14 about here

- - - - - - - - - - - - - - - - - - -

It appears, then, that the search for acoustic invariances characterizing each of the vowel phonemes has met with only qualified success when proceeding along the route of spectrographic measurement of vowel parameters, notably the frequencies of the first and second formants. Average parameter values may be computed, of course, and different average values may be associated with each of the vowels, but the confidence intervals around these means are exceedingly large. Since each vowel is associated with a wide range of parameter values, some investigators have used the discriminative behavior of listeners as a guide to which restricted sets of values may be characteristically associated with the vowel. The basic strategy is to sample the vowel space and then present these samples to listeners for identification; the experimental procedures are of essentially three types:

(1) A large set of natural vowels is collected, so that a broad distribution of phones in the vowel space may be obtained. This technique has the disadvantages that vowel parameters cannot be manipulated independently and the vowel space cannot be sampled systematically. It is superior to techniques (2) and (3), however, in that the stimuli to be judged have the redundancies and irregularities of natural language and hence the results are more generally applicable.

(2) A set of natural vowels is distorted by filtering, masking and the like so that certain cues are attenuated or obliterated entirely. When the degree of distortion begins

to yield a reduction in intelligibility, the range of significant parameter values is indicated.

(3) A set of vowels is synthesized electronically. This technique has the advantages that vowel parameters may be varied separately and systematically over the vowel space but its results need to be confirmed by experiments with natural language.

Illustrative of the first approach is an experiment by Fairbanks and Grubb (1961) in which 126 isolated vowel phones were collected from seven speakers. The phones corresponding to each phoneme were sorted into three categories: (a) those approved by the speaker upon playback as representative of the vowel he intended, (b) those self-approved phones which were also identified correctly by three fourths or more of a panel of listeners, and (c) the four phones in class (b) that received the highest ratings of "representativeness" by the panel of listeners. Figure 15 shows the way in which these three screening variables converge on characteristic parameter values for each of the vowels. These values correspond well with the average values obtained by spectrographic measurement, (Table I).

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 15 about here

- - - - - - - - - - - - - - - - - - - -

Illustrative of the second approach is an experiment by Carterette and Møller (1962) on the identification of Swedish vowels after very sharp high-pass or low-pass filtering. Figure 16 compares their results with the outcome of two similar experiments on sustained English vowels. As the cutoff frequency of the low-pass filter is increased and more of the vowel spectrum becomes audible, correct recognition increases monotonically. Similarly, as the cutoff frequency of the high-pass filter is decreased, recognition improves. The point of intersection of the two functions indicates that the part of the vowel spectrum below 1,000 cps makes as much a contribution to identification as the part above this frequency. The agreement among the three experiments on the point of inter-

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 16 about here

- - - - - - - - - - - - - - - - - - - -

section is particularly noteworthy in view of the differences among them in the context and durations of the vowels employed and the language and training of speakers and listeners. Carterette and Møller showed, moreover, that the discontinuities in the functions in Fig. 16 reflect the non-uniform distribution of formant frequencies in the vowels. They found that the form of the functions could be predicted on the assumption that the percent correct identification is proportional to the weighted sums of the number of formants available at each value of filtering.

The parameter values that are critical to the identification of each of the vowels are revealed by considering the identification functions and spectra of each of the vowels separately. For example, Lehiste and Peterson (1959a) report that for the vowel [i] high-pass filtering above 550 cps effectively removed F1 and preserved identification at 80 per cent, whereas low-pass filtering at that frequency destroyed recognition entirely. (Nicely enough, the one formant preserved under the latter condition was identified as [u]). Total accuracy of recognition was obtained if filtering preserved either F1 and F2, or F2 and F3. Figure 17 shows the shifts in vowel identification that were observed by Miller (1956) when he rejected by filtering all but the first formants of the vowels employed by Peterson and Barney (Table I). By deleting all formants above the first, the vowel space of Fig. 17 is effectively collapsed onto its ordinate. It might be expected, therefore, that the lines of confusion would run horizontally through the vowel space, and so they do, but with certain noteworthy exceptions which will be discussed later.

- - - - - - - - - - - - - - - - - - -

Insert Fig. 17 about here

- - - - - - - - - - - - - - - - - - -

The third approach to finding acoustic invariances for the vowels based on the discriminative behavior of listeners is illustrated by an experiment with vowels synthesized by the conversion of hand-painted spectrographic patterns into sound. The conversion device, called pattern playback, is schematized in Fig. 18 and its operation is summarized in the caption. Delattre and his collaborators (1952) synthesized 235

- - - - - - - - - - - - - - - - - - -

Insert Fig. 18 about here

- - - - - - - - - - - - - - - - - - -

two-formant vowels on pattern playback, and presented to listeners for identification those which they judged closest to the 16 cardinal vowels of the I.P.A. It should be noted that the vowels were presented in isolation and all had the same fundamental frequency, sound pressure level and duration. The absence of these intrinsic differentia, described below, might be expected to reduce the accuracy of vowel recognition. Figures 19 and 20 represent the synthetic patterns in an acoustic vowel diagram and also show the accuracy of identification of each pattern. In listening test C, only the

- - - - - - - - - - - - - - - - - - -

Insert Fig. 19 about here

- - - - - - - - - - - - - - - - - - -

seven widely separated patterns were presented; the accuracy of recognition ranged from 75 to 100 per cent. Except on one occasion, each of the patterns was always identified as the indicated vowel or its adjacent neighbor. When all 16 patterns were presented, in listening test A (Fig. 20), only the egregious [i] retained its high level of intelligibility. The patterns corresponding to the two vowel phonemes that do not occur in

- - - - - - - - - - - - - - - - - - -

Insert Fig. 20 about here

- - - - - - - - - - - - - - - - - - -

English, [ɯ] and [ɤ], had the lowest recognition scores of all the patterns but one.

An experiment by Fry _et al_. (1962) permits a closer look at the pattern of vowel identifications when the frequency parameters of two-formant patterns are varied systematically. Imagine a straight line drawn through the acoustic vowel diagram in Fig. 3 so that it intersected the regions associated with [ɪ], [ɛ], and [æ]. This line was sampled at thirteen points, and the corresponding two-formant patterns were synthesized with the formant frequencies shown in Table III. When the patterns were presented repeatedly,

- - - - - - - - - - - - - - - - - - -

Insert Table III about here

- - - - - - - - - - - - - - - - - - -

in mixed order, to a group of listeners, the identification gradients shown in Fig. 21

were obtained. With the synthetic vowels employed, it was possible to obtain consistent

Insert Fig. 21 about here

identifications for [I] and [æ], but the parameter values that evoked the most [ɛ] responses also yielded some [I] and more [æ] identifications. (The overlap of [ɛ] and [æ] in natural language was discussed in connection with Fig. 7 above.) Unfortunately, the acoustic vowel space has not been charted systematically in the manner of this experiment. It would be most interesting to examine the set of identification surfaces that would be obtained and to determine the effects on their form of various psychological transformations on the coordinates of the vowel space. Recent advances in techniques of speech processing make it possible to perform this experiment with natural vowels whose harmonic components are systematically altered in frequency and relative amplitude.

The limited findings available from studies employing synthetic vowels to delimit the ranges of their parameter values that have a discriminative function confirm the conclusions reached on the basis of spectrographic measurements presented earlier. Despite a wide range of parameter values in natural language, the vowels of English may be specified in terms of two formants whose frequencies are below 3,000 cps. It must be acknowledged that this one generalization does not go very far in explaining the relation between the acoustic and discriminative properties of vowels in the stream of speech. Since the two-formant vowel space below 3,000 cps shows great variability and overlap of vowels, whereas identifications of these vowels are highly consistent and non-overlapping, it seems likely that the listener utilizes at least three kinds of additional cues in identifying vowels in the stream of speech: (1) those cues provided by the immediately preceding or following segments of speech; (2) those cues provided by the non-linguistic environment; (3) those cues provided by acoustic properties of the vowel other than its formant frequencies. The first two considerations are beyond the scope of this survey. The acoustic parameters of the vowels other than formant frequencies include fundamental frequency, the bandwidths and relative amplitudes of the formants, and the amplitude and duration of the total vowel. We now turn to research findings concerning these parameters of the vowels.

Fundamental frequency

Inspection of the average fundamental frequencies of the vowels, shown in Table I, reveals that each vowel has a characteristic fundamental frequency which varies directly with the tongue height associated with the vowel (Cf. Fig. 3 and House and Fairbanks, 1953). These measurements are plotted in Fig. 22 along with more recent data for a

- -
Insert Fig. 22 about here
- -

single speaker from Lehiste and Peterson (1961a) which confirm the prior results and supplement them for the diphthongs.[4] A detailed examination of the effects of the consonantal environment on the fundamental frequency of the intervening vowel or diphthong showed that the final consonant had no systematic effect whereas the initial consonant had a marked effect: the fundamental frequency was higher after a voiceless consonant and lower after a voiced consonant. This confirms the earlier findings of House and Fairbanks (1953) to the same effect.

The preceding fings apply to general American English where fundamental frequency is not the primary cue - if indeed it is a cue at all - in discriminating among the vowels. It is instructive to consider the vowels of a tone language where the inter- and intra- segmental variations in fundamental frequency are marked and serve to differentiate utterances in minimally contrasting pairs, just as vowel formant frequencies distinguish minimal pairs of words in English. For example, Fig. 23 presents spectrograms of two series of Thai monosyllables:

- -
Insert Fig. 23 about here
- -

naa (midtone, 'field'); naà (low tone, 'custard apple'); naá (high tone, 'mother's younger sibling'); naâ (falling tone, 'face'); naǎ (rising tone, 'thick') and k'un ('you'); k'ùn ('muddied'); k'ún ('familiar'); k'ǔn ('to feed').

A summary picture of the five tones in Thai and their range of variation is presented in Fig. 24.

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 24 about here
- - - - - - - - - - - - - - - - - - - -

Formant bandwidths

House and Stevens (1958) measured the impulse response of the vocal tract by recording the transient produced by a spark gap inserted in the mouth while the articulators were set to phonate one of eight American English vowels. The damped oscillations (formants) that together make up the complex transient response of the tract, were sorted by filtering and the half-power bandwidth of each formant was deduced from the decay time of its corresponding transient. With the glottis and velum closed, the average bandwidths for formants one, two, and three were, respectively, 54, 65, and 70 cps. This index of the damping of the vocal resonances reflects energy losses primarily from radiation from the mouth opening and absorption of sound by the walls of the vocal tract. Sufficient data are not available to describe the vowels differentially according to their formant bandwidths, but the measurements at hand show that differences among the vowels of English are not large in this respect.

As indicated earlier, listening tests are employed to complement acoustic measurements in an effort to characterize the vowels in terms of a small portion of the range over which their parameters may vary. In the case of formant bandwidth, House (1960) used an electrical synthesizer to generate five vowels whose parameter values corresponded to those reported for natural vowels by Peterson and Barney (Fig. 8). He then varied the formant bandwidths of each of the vowels over seven sets of values which ranged from narrow (Set A consisted of 40, 50, 60 cps) to wide (Set G, 220, 260, 300 cps). When this array of vowels was presented in paired comparisons to listeners, the preference functions shown in Fig. 25 were obtained. Overall, the most preferred synthetic vowels have

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 25 about here
- - - - - - - - - - - - - - - - - - - -

narrow bandwidths, ranging between Set A and Set B (70, 85, 100 cps). These values agree well with the acoustic measurements on natural language reported above. In particular, it will be seen from the Figure that i and u have the smallest preferred bandwidths, with [ɛ] and ɔ next in line and [æ] at the bottom of the rank order. This ordering of the vowels according to bandwidth corresponds to that according to tongue height (Cf. Fig. 3) and may reflect the correlation between the height of the tongue and that of the velum in phonation, which has been reported previously. A lowered velum, which couples the nasal cavity to the vocal tract, adds another source of acoustic loss, thus increasing the damping of the resonances and hence the formant bandwidths.[5]

Formant amplitudes

Average formant amplitudes for general American vowels appear in Table I.[6] In order to assess the role of formant amplitude in the identification of the vowels, the experiment employing vowels synthesized on pattern playback, discussed above, included some systematic manipulations of the amplitudes of F1 and F2. Reference to Fig. 20 will assist in interpreting the results. For those vowels with widely spaced formants, progressive reductions in the level of F1 led finally to the destruction of vowel quality. For the back vowels, whose formants are relatively close together, however, the same operation caused identification to shift to that for the adjacent vowel: [u] and [o] "moved" to [ɔ], [ɔ] to [ɑ] and [ɑ] to [æ]. A reasonable inference from these results is that the listener effectively combines the levels of F1 and F2, perhaps by a process like averaging them, so that a reduction in the level of F1, which yields an increase in the relative contribution of F2, raises the "mean" formant with a corresponding vertical shift in identification. This hypothesis finds some support in additional experiments by Delattre et al. (1952) in which the back vowels were successfully synthesized by a single formant. When the level of the second formant was progressively lowered in the 16 synthetic vowels, corresponding shifts occurred, now moving from left to right across the vowel diagram in Fig. 20; this result is in accord with that obtained by low-pass filtering of the vowels, as shown in Fig. 17. The horizontal shifts in identification may be interpreted like the vertical shifts, in terms of the averaging hypothesis described by Delattre and his collaborators. Complementing these findings, Miller (1953) showed that F 2 and F 3 in the front vowels combine in their effects on identification and that increasing the levels of F 2 and F 3 has an effect similar to decreasing the level of F1.

Vowel amplitude

The amplitudes of isolated, sustained vowels and of vowels in consonantal context have been reported by several experimenters (Fairbanks, House and Stevens, 1950; House and Fairbanks, 1953); Fig. 26 shows the measurements obtained under the two conditions in a recent study by Lehiste and Peterson (1959b) whose findings are representative.

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 26 about here
- - - - - - - - - - - - - - - - - - - -

It will be seen that when the several vowels are spoken with approximately equal effort, the differences in phonetic quality are associated with "intrinsically" different amplitudes; this is the case whether isolated vowels or syllable nuclei (which were about 3 db weaker) are considered. It will be necessary to bear these baseline differences in mind when the amplitude changes that signal stress are discussed below.

The effects of the initial and final consonant on the amplitude of the syllable nucleus are shown in Table IV. As observed with the vowel parameter of fundamental frequency, the major effect is associated with the voicing property of the consonants: the voiced member of each consonant pair appears higher in the rank order of Table IV than does its voiceless counterpart. These findings are in agreement with those of House and Fairbanks (1953), who suggest that continuation of phonation throughout the consonants as well as the vowel would be likely to favor greater maximum intensity in the vowel segment.

- -
Insert Table IV about here
- -

Vowel Duration

The durations of syllable nuclei in English, averaged over their occurences in consonantal environments, are shown in Fig. 27. It seems possible to distinguish

- -

Insert Fig. 27 about here

- -

intrinsically short syllable nuclei, the "lax" vowels [I, ɛ, ɔ, ʊ] from among the rest. When vowel durations are examined with respect to the preceding and following consonants, it is found that the initial consonant has little effect on the duration parameter, whereas the final consonant has a marked effect, as illustrated in Fig. 28. Both intrinsically short (bottom curve) and long vowels are prolonged before voiced consonants,

- -

Insert Fig. 28 about here

- -

as compared with their voiceless counterparts. Sharf (1962) found the ratio of the average durations of vowels before voiced and voiceless plosives to be 3/2. With voicing constant, vowels also seem to be longer before fricative than before stop consonants. These results agree with those of House and Fairbanks (1953) and Denes (1955); the latter study showed that the perception of a final consonant as voiced increases as the duration of the preceding vowel is increased.

In a more recent study, House (1961) confirmed the preceding findings and summarized the effects of the three variables cited plus one more: the tongue-height associated phonetically with the vowel. This summary is shown in Fig. 29 where the tense/lax opposition in the vowels and the voiced/voiceless opposition in the following consonant

- -

Insert Fig. 29 about here

- -

are identified as primary variables affecting vowel duration. House weighs the considerations, among them cross-linguistic findings, which indicate that the primary influences are learned by speakers of English, whereas the secondary influences are a function of articulatory variables. Delattre (1962a) puts forth additional arguments for acknowledging both physiological and phonemic determinants of vowel duration in English.

With the preceding findings in mind we may finish unraveling the pattern of vowel confusions shown in Fig. 17. The vowels at the center of the formant space, [ɪ], [ɛ], [ʌ], [ʊ], are short vowels and hence discriminable from the remaining vowels on the basis of duration cues alone. This explains the absense of confusions between [æ] - [ʌ], [ʌ] - [ɔ] and [ɛ] - [ɔ], despite the similarity of their F1 frequencies and hence of their spectra after filtering. The data shown in Fig. 17 largely accord with the notion that there are two quadrilaterals in the acoustic vowel space: one for the short (lax) vowels which lies inside of one for the long (tense) vowels.

Once again, it is instructive to examine a language in which the vowel parameter serves systematically to differentiate among utterances. Returning to Thai for our examples, Fig. 30 presents spectrographic analyses of the sentences "I'm saying the word 'goat' for you to listen to" and "I'm saying the word 'to be defeated' for you to listen to." There are many such minimal pairs in Thai: sin 'riches', siin 'moral code'; het 'mushroom', heet 'reason', which must make fertile soil for comedians and punsters. Abramson (1962) lists 14 pairs in all, with the durations of the long vowels about two and a half times that of their short counterparts.

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 30 about here
- - - - - - - - - - - - - - - - - - - -

Nasalization of vowels

The effects on the parameters of oral vowels of coupling the nasal cavities to the vocal tract are complex and the results of several recent investigations are not entirely consonant.[7] A recurrent and major consequence of nasalization is an increase in the bandwidth of F1 resulting from an additional energy loss in the nasal cavities, as discussed earlier. This increase in the bandwidth of the first formant is concomitant with a decrease in its amplitude; both acoustic parameters are a function of the size of the cavity which is coupled to the vocal tract. The relations among the three variables are illustrated by the set of spectrum envelopes for the vowel u shown in

Fig. 31. The different spectra were obtained by varying the parameters of an electrical

- - - - - - - - - - - - - - - - - -
Insert Fig. 31 about here
- - - - - - - - - - - - - - - - - -

analog of the vocal tract in a way that corresponded to increasing the area of a sidebranch to the tract. When this area is increased from 0 to 4 cm^2 a marked increase in the bandwidth of F1 and a decrease in its level are observed. A corresponding increase is obtained in the perception of nasality, as shown in Fig. 32. Nasalization is recog-

- - - - - - - - - - - - - - - - - - -
Insert Fig. 32 about here
- - - - - - - - - - - - - - - - - - -

nized as a distinctive feature in many languages but is generally not considered such in English. Malecot (1960b) has demonstrated, however, that in words such as "camp", "hint", "bunk" and the like, a nasal segment is not to be found acoustically, perhaps because of the sound interval necessary as a cue for the voicelessness of the following stop. As a result, it is the nasalization of the vowel nucleus which differentiates "camp" from "cap", "hint" from "hit", "bunk" from "buck", etc.

The plosives

Spectrograms of the plosives [p, t, k, b, d, g] , spoken in initial position before [a] by a Russian informant, appear in Fig. 33. Accompanying each spectrogram is an amplitude section taken at the indicated point in time. The graphs have been arranged in columns according to the traditional phonetic classification by "place of

- - - - - - - - - - - - - - - - - - -
Insert Fig. 33 about here
- - - - - - - - - - - - - - - - - - -

articulation": bilabials [p, b] , alveolars [t, k] and velars [d, g]; they have been arranged in rows according to "manner of articulation": voiceless [p, t, k] and voiced [b, d, g]. Acoustic-phonetic research on the consonants, as on the vowels, has involved both spectrographic measurements of natural language and listening tests with altered utterances or synthetic patterns. In contrast to research on the vowels, however, experiments with the consonants have most often employed synthesized speech, perhaps for the very reason that the relevant acoustic properties of the consonants appear more complex.

When we examine the spectrogram of [da], for example, several acoustic features suggest themselves as possible cues participating in the distinction between [d] and the other five plosives.

(1) There is a brief burst of energy, at the point marked "I"; the amplitude section reveals the locus of the burst more clearly, approximately 1500 cps.

(2) There is a marked positive transition of F2 which begins above the formant frequencies of the following vowel.

(3) The transition of F3 appears to be positive, although this transition is not as clearly marked as those for F1 and F2.

(4) There is a marked, negative transition of F1.

(5) A lengthy "voice bar" is observed which reflects the vibration of the vocal cords prior to the explosion of the voiced plosive.

When the spectrogram for [da] is compared with those for the other five plosives there is an indication that all these acoustic properties may serve as cues for discriminations among the plosives, since the properties clearly vary - with the exception of (3) - from one plosive to the next and there is initial evidence that they vary systematically. In an inspection of Fig. 33, the voice bar is the most obvious feature distinguishing two groups of plosives according to manner of articulation. It will also be observed that the F1 transition is considerably less steep and has a later onset in the voiceless member of each pair of plosives. Another acoustic property of the voiceless as opposed to the voiced plosives is the presence of weak noise rather than harmonics in the beginning parts of the formant transitions. This property, which corresponds to the phonetic feature of aspiration, is difficult to detect in Fig. 33, although the amplitude section at the point marked III for [ka] reveals its presence. The feature of aspiration in the voiceless plosives can be more clearly seen in the spectrogram of [kʌp] in Fig. 34.

- -

Insert Fig. 34 about here

- -

With respect to place of articulation, the bursts and the formant transitions are candidates for the role of discriminative cues. The amplitude spectra of the bursts show a maximum at low frequencies for the bilabials, a more spread and even distribution for the alveolars, and for the velars a maximum again, this time at a frequency somewhat above that for the bilabials. The transition of the second formant is negative in the case of the bilabials and positive otherwise, with some indication in these data that it is more steep in velars than in the alveolars. Thus, the set of acoustic cues that distinguishes the plosives according to manner of articulation may be different from the set which distinguishes them according to place of articulation.

As observed earlier, experiments with altered and synthetic speech are required to delimit a range of values for each of these parameters that may characterize the plosives. In considering the findings of these experiments, we will first ask what are the acoustic cues which distinguish the plosives according to place of articulation and then what are the cues which distinguish between them according to manner of articulation. This strategy is based not only on traditional phonetic distinctions and the corresponding sets of acoustic features which we have just noted, but also on some evidence from a classic study by Miller and Nicely (1955) which indicates that "place" cues and "manner" cues are fundamentally different. These investigators presented utterances, each composed of one of 16 English consonants followed by [a], to a group of subjects who listened over a noisy channel and attempted to write down what they heard. The stimulus-response pairs (confusions and correct identifications) were examined with the consonants sorted into groups in several ways. The two tabulations of present concern are the percent confusions between voiced and voiceless consonants and between front, mid, and back consonants. Miller and Nicely expressed the relative accuracy of identification in terms of the information transmitted at each of six signal-to-noise ratios, as shown in Fig. 35. The cues associated with the voicing

- -

Insert Fig. 35 about here

- -

of the plosives (and fricatives and nasals)--that is, with their manner of articulation--are more resistant to noise at all levels than the cues associated with their place of articulation. Moreover, in an analysis of the total confusions in recognition, these two classes of cues seemed to make relatively independent contributions to the total information transmitted.

Place of articulation [p, b: t, d: k, g]

(1) The burst.

First we may note that systematic spectral measurements of the bursts of English plosives confirm our initial impressions based on Fig. 33. According to Halle, Hughes and Radley (1957), [p, b] have a primary concentration of energy in the low frequencies (500-1500 cps), [t, d] are characterized by a flat spectrum or one in which the higher frequencies predominate and [k, g] show concentrations of energy at the intermediate frequencies (1500-4000 cps). These investigators noted, in addition, that the burst frequency depends on the following vowel - especially in the case of the plosives [k, g] where the frequency was found to be higher before front vowels and lower before back vowels.

Liberman, Delattre and Cooper (1952) set out to determine the role of the burst frequency and of the formant frequencies of the following vowel in the identification of the voiceless plosives. The sequence of plosive followed by vowel can be synthesized on pattern playback with hand-painted spectrograms like that shown in Fig. 36C (cf. [kɑ],

- -

Insert Fig. 36 about here

- -

Fig. 34). The array of burst and formant frequencies employed as variables in the experiment are shown in Figs. 36A and B, respectively. The identifications of the synthetic patterns by listeners as [p], [t] or [k] are summarized in Fig. 37. Largely

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 37 about here
- - - - - - - - - - - - - - - - - - - -

in accord with spectral measurements of the plosives in natural language, the optimal burst for [p] is at lower frequencies, that for [t] is at higher frequencies, and that for [k] at intermediate frequencies but varying according to the second formant (or fronting) of the following vowel. The latter effect is particularly striking when the burst frequency is, for example, at 1440 cps. Before [i], this schematic stop was called [p] 59 times out of 60; from [i] to [a] there was a progressive drop in [p] (increase in [k]) judgments, terminating with the response [k] 47 times in 60 presentations; from [a] to [u], [k] judgments decreased as those for [p] increased, until at [u] the plosive was called [p] 42 times in 60. This complex acoustic invariance--identical bursts identified as different consonants, and different bursts identified as the same consonant--has been considered by some scientists as evidence that the syllable is the smallest acoustic unit in speech, while others consider it evidence that the invariances distinguishing consonants in speech perception are basically articulatory rather than acoustic.[8]

The preceding findings on burst frequencies for the voiceless plosives have been extended by the same techniques of speech synthesis to the voiced plosives (Hoffman, 1958) and they have been confirmed by experiments with natural language. Malecot (1958) demonstrated that if the release of a postvocalic voiced plosive is spliced onto the end of an isolated vowel (after a closure interval of 100 msec), the plosive will be accurately identified, despite the absence of formant transitions. This finding is in accord with the results of the experiment by Liberman, Delattre and Cooper (1952) described above (cf. Figs. 36, 37). Moreover, Malecot found that when the releases of [ɛb, ɛd, ɛg] and [ɔb, ɔd, ɔg] were interchanged in all possible combinations, the product typically was identified as the plosive from which the release was taken rather than as the plosive appropriate to the formant transitions preceding the release. Essentially the same results were obtained for the voiceless plosives in this experiment and for all six English plosives in a subsequent study by Wang (1959). These findings place special emphasis on the importance of the burst as a cue to the place of articulation of plosives.

Schatz (1954) confirmed with stimuli prepared from natural language that burst frequency and formant frequency interact in their control of plosive identification. The utterances "keep, cop, coop" and "heap, hop, hoop" were recorded several times each on magnetic tape and each of the plosives in the first set was spliced next to each of the vowels in the second set. According to the findings with synthetic speech (Fig. 37), a burst that sounds like [k] before [i], as in "keep", sounds like [t] before [a] and like [p] before [u] --and this is exactly what Schatz found. The correspondence between the two studies is summarized in Table V.

- - - - - - - - - - - - - - - - - - -

Insert Table V about here

- - - - - - - - - - - - - - - - - - -

(2) The transition of F2.

The role of the transition of the second formant, which follows the burst of a prevocalic plosive, has also been studied by means of speech synthesis with pattern playback. The two-formant synthetic vowles which were employed are shown in Fig. 38C. The varying transitions of F2 adjoined to each of the vowels are shown in part A of the figure; part B shows a typical pattern which was presented for identification as [p] , [t] or [k] with one group of listeners and as [b], [d] or [g] with another group.

- - - - - - - - - - - - - - - - - - -

Insert Fig. 38 about here

- - - - - - - - - - - - - - - - - - -

The distributions of identifications were similar for the two groups[9] and are summarized by the contour plot for [b, d, g] shown in Fig. 39. Confirming and extending our observations in Fig. 33, we find that [b, p] are predominantly associated with negative

- - - - - - - - - - - - - - - - - - -

Insert Fig. 39 about here

- - - - - - - - - - - - - - - - - - -

F2 transitions while those for the remaining plosives are positive. Of these, it is noteworthy that the transitions for [k, g] vary according to the frequency of F2 (or fronting) of the following vowel: higher positive transitions before front vowels, lower before back vowels. In the absence of the burst cue, [d, t] before [i] and [g,k] before [a] were not synthesized with appreciable success by any value of the F2 transition. The findings indicate that F2 transitions can be important cues for distinguishing among places of articulation both for the voiced and voiceless plosives but the rules for their invariance seem quite complex.

The F2 transitions that lead to the most reliable identifications of each of the plosives before the several vowels are arranged in sets in Fig. 40 in order to show graphically an important invariance in their acoustic properties. Although the same plosive requires differing F2 transitions for its synthesis before different vowels, it appears from the arrangement in Fig. 40 that the transitions associated with each of the plosives have something in common, namely a point of intersection or locus. At the left of the Figure, the F2 transitions of the bilabials are extended to a locus at 700 cps. Next, the locus of the alveolars is shown at 1800 cps. Finally, the locus of the velars is traced to 3000 cps, with the exception of the transitions before the back vowels [ɔ, o, u], whose anamolous behavior was noted in connection with Fig. 39.[10] The identification of the loci was also determined in part by the results of an experiment in which 13 first formants with different frequencies and transitions were variously combined with a like number of second formants which had differing frequencies but no transitions (Delattre, Liberman & Cooper, 1955). As indicated at the right of Fig. 40, the frequencies of the straight second formants at which the optimal plosives were heard correspond to the respective loci.

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 40 about here

- - - - - - - - - - - - - - - - - - - -

Associated with each locus and place of articulation is, of course, a configuration of the vocal tract, indicated in Fig. 40 by a tracing from a cineradiogram. The relation between the acoustic-phonetic definition of locus and the corresponding configuration of the vocal tract has been described in this way by Delattre, Liberman and Cooper (1955):

"The second-formant transitions rather directly represent the articulatory movements from the place of production of the consonant to the position for the following vowel. Since the articulatory place of production of each consonant is, for the most part, fixed, we might expect to find that there is correspondingly a fixed prequency position - or "locus" - for its second formant" (p.769).[11]

Qualified support for the general applicability of the concept of locus has come from spectrographic measurements of natural language. Green (1958) prepared tracings of the F2 transitions of utterances like [bi: bi: b] and then extended convergent tracings for each consonant to a point of intersection. Many of the transitions did not converge and, in the case of [k, g], the formant transitions were, on the contrary, divergent. In those cases where a locus could be established, however, the average frequencies of the loci compared favorably with those arrived at by speech synthesis.[12] Another point of agreement relates to the time interval between the locus and the onset of the F2 transition. The samples of natural speech shown in Fig. 33 and of synthetic formants in Fig. 40 make it clear that the formant transitions do not actually begin at the locus; there is a delay before the onset of the transitions which may be taken as an indication that the articulators have partly moved away from the point of production before voicing begins. Experiments with synthetic speech (Delattre, Liberman & Cooper, 1955) showed that the optimal delay for plosive identification was approximately 50 msec. Measurements on natural language (Green, 1958) yielded an average delay of 60 msec between the locus and the onset of the F2 transition.

Two experiments with natural language (Householder, 1956; Andressen, 1960) confirm that formant transitions alone are sufficient to distinguish among the voiceless plosives, in the absence of cues provided by their release (burst and aspiration). When a series of words such as "lip, lit, lick" was spoken with unreleased final plosives (this seems to be accomplished by clamping the glottis shut at the right moment) identifications were reasonably accurate under good listening conditions. It is noteworthy, however, that if a back vowel was the syllable nucleus, the accuracy of [k] identifications was seriously impaired. These findings are in accord with the outcome of research utilizing synthetic speech. Formant transitions alone may serve as a cue for the place of articulation of the plosives, but the presence of a burst takes on particular importance with the velar plosives, especially when adjacent to a back vowel.

(3) The transition of F3

Two experiments have employed three-formant patterns, systematically varying the transitions of F2 and F3 (but not their formant-frequencies), in order to assess the contribution of the third formant to the identification of plosives (Harris et al., 1958; Hoffman, 1958). Both studies found that identification of the plosives in the absence of F3 was unanimous at some value of F2 transition, so the positive contribution of F3 transitions could not be fully assessed. An exception to this generalization was the poor identification of [d] before [i] without F3, observed by Harris et al. (1958); in this case, the most positive transition of F3 employed (+4) gave the best results. Even when identification reached 100 per cent at some transition of F2, in the absence of F3, the effect of adding F3 could be detected as a change in the relative distributions of the identification responses. With respect to [d], the distribution of identification responses increased in area (i.e., there were more [d] responses to more values of F2 transition) when more positive F3 transitions were employed. Exactly the opposite was true for [b] and [g].

The available evidence indicates, then, that the optimal F3 transition is positive for the alveolars and negative for the bilabials and velars. There is also some evidence from the two studies we have been discussing that the interaction effects between burst frequency, F2 transition and F3 transition are relatively small. Hoffman (1958) found that the addition of various F3 transitions to patterns containing various burst frequencies and F2 transitions changed identifications in the same way as when F3 was added to the series of patterns which did not contain bursts. Similarly, when bursts were added to the three-formant patterns, identifications changed in the same way as when bursts were added to the two-formant patterns. Thus, extending Miller and Nicely's observation that the acoustic cues associated with manner of articulation are independent of those for place, we have Hoffman's observation that the several acoustic cues associated with place also make independent contributions to identification.

Manner of articulation [p, t, k: b, d, g]

In the light of the preceding findings on the acoustic properties that characterize prevocalic plosives according to their place of articulation, Liberman, Delattre and Cooper (1958) prepared the highly schematic patterns shown in Fig. 41, which successfully synthesized the plosive-vowel combinations indicated. In order to determine which acoustic properties of these voiced plosives play a role in distinguishing them from their voiceless counterparts, a series of patterns was prepared for each of the nine plosive-vowel syllables in Fig. 41, in which the F1 transition was successively cut back (first deleting the voice bar) in steps of 10 msec. The series based on the spectrographic pattern that synthesized [bæ] is shown in Fig. 42. Nine series of these three-formant patterns were presented in random order to listeners, who were instructed to identify each pattern as one of the plosives. With the voice bar present and no cutback in F1, all the patterns were identified unanimously in accord with the labels shown in Fig. 41. The deletion of the voice bar reduced the number of [b] and increased the number of [p] responses but had virtually no effect on voicing judgments for the alveolar or velar plosives. With

Insert Fig. 41 about here

Insert Fig. 42 about here

increasing amounts of cutback, the identification of each voiced plosive increasinly shifted to its voiceless counterpart, with the bilabials yielding first, the velars second and alveolars third. These findings indicate that, in distinguishing the plosives according to manner of articulation, the transition of F1 is a major cue whereas the voice bar plays a secondary role (also see Fischer-Jørgensen, 1954). Precisely which properties of the F1 transition were responsible for the voiced-voiceless distinction remained unclear, however, since this experiment confounded changes in the relative time of onset of F1 with changes in its frequency at onset. Therefore, the series of patterns shown in Fig. 43 was prepared and presented to listeners. The pattern at the left, with 0 msec cutback of F1, generates the syllable [do]. This syllable was chosen for synthesis because the low F1 frequency in the vowel permits synthesis of the plosive without an F1 transition. Therefore, if increasing delays in the onset of F1 cause identifications to shift from [d] to [t], the relative time of onset of F1 is isolated as a cue to voicing. And this is exactly what happened; [d] judgments increasingly gave way to [t] until, with delays greater than 25 msec, [t] identifications were predominant.

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 43 about here

- - - - - - - - - - - - - - - - - - - -

Thus, three of the four acoustic properties which differentiated the voiced and voiceless plosives in the spectrograms of Fig. 33 have proven to play a role in the discrimination of voicing. The fourth property, the aspiration following the burst of a voiceless plosive (cf. Fig. 34), did not escape the paint brush of Liberman, Delattre and Cooper. The series of patterns shown in Fig. 42 was prepared once again for each of the syllables shown in Fig. 41, but this time the transitions of F2 and F3 were "filled with noise" (that is, speckled like the schematic bursts in the Figure) up to the time at which the transition of F1 began. In this respect, the patterns resembled the spectrogram in Fig. 34. We have seen earlier that increasing delays in the onset of F1 led to a shift in identifications from voiced to voiceless plosives. When the aspiration cue was now added, the shift from voiced to voiceless identifications took place at shorter values of the delay in the onset of F1.[13]

Further evidence of the importance of aspiration as a cue to voicelessness comes from an experiment with natural language. Reeds and Wang (1961) tape-recorded English words that begin with [s] followed by a plosive (in this position, plosives are neither voiced nor aspirated in English) and then they cut away the segments of tape on which the friction of [s] was recorded. Words such as "pat, tore, cold" were extracted in this way from recorded utterances of "spat, store, scold,", etc. and then presented to listeners for identification as "pat" or "bat", "tore" or "door", "cold" or "gold". In virtually every case, the unaspirated voiceless plosives of the truncated syllables were identified as voiced stops.[14]

We may conclude, then, that all the acoustic differentia which are apparent on spectrograms of the voiced and voiceless plosives play a role in distinguishing between them according to their manner of articulation.[15]

The fricatives and affricates

Spectrograms of English fricatives and affricates, spoken by a native midwest American, appear in Figs. 44-46. Once again, the consonants are grouped by place and manner of articulation. Figure 44 shows the front group of fricatives, voiceless

- - - - - - - - - - - - - - - - - -

Insert Figs. 44-46 about here

- - - - - - - - - - - - - - - - - -

f θ and (underneath) voiced [v ð], as well as the midgroup, voiceless s and voiced z. Figure 45 shows the back group, voiceless [,h] and (underneath) voiced [ʒ, ɦ]. The affricates are shown in Fig. 46; voiceless [tʃ] and (underneath) voiced [dʒ]. It must be borne in mind, however, that front-mid-back are approximate categories: whereas the "front" stop consonants [p, b] are bilabial, [f, v] are labiodental and [θ, ð] are dental; in the mid group, [t, d] are alveolar, as are [s, z] and the affricates [tʃ, dʒ], but [ʃ, ʒ] are palatal; finally, in the back group [k, g] are velar, but [h, ɦ] are glottal.

The tripartite grouping of the fricatives is based on their points of articulation and on their corresponding acoustic properties. The spectrograms of Figs. 44-46 reveal three acoustic parameters of the fricatives that may play a role in distinguishing among them according to place of articulation: the frequencies of the fricative formants, the intensity of the fricative and the formant transitions of adjacent vowels.

Representative spectra of three voiceless fricatives are shown in Fig. 47. The dotted curves were obtained by measurements of English words; the solid curves were obtained by measuring the output of an electrical analog of the vocal tract whose parameters were adjusted to yield two resonances and one antiresonance (Hughes & Halle, 1956). It will be observed that the frequency of the first major resonance of [f] is above 7,000 cps, while that for [s] is lower, typically 5,000 cps, and the first resonance of [ʃ] lower still, about 2500 cps (this resonance is low enough to appear on the spectrogram in Fig. 45). The voiced counterparts of each of these fricatives have comparable spectral distributions.[16]

- -

Insert Fig. 47 about here

- -

In order to establish that these spectral differences make a difference, that is, exert differential control over the discriminative behavior of English listeners, Heinz and Stevens (1961) synthesized isolated voiceless fricatives with an electrical analog of the vocal tract which provided one resonance and one antiresonance applied to noise excitation. When the resonance was varied from 8,000 to 2,000 cps (with the antiresonance always one octave below), listener identifications varied from [f] and [θ] through [s] to [ʃ]; peaks in the identification functions were located near peaks in the spectral distributions.[17] We may conclude that the resonant frequency of a fricative is a major cue for its identification. It is noteworthy, however, that the listeners failed to discriminate [f] and [θ] when only spectral cues in the fricative segment were available. This finding suggests that differences in intensity or in the transitions of vowel formants play a role in distinguishing among some of the fricatives. It is doubtful that "intrinsic" differences in intensity have a major discriminative function, since the sound pressure level of a speech segment is subject to many concurrent variables, phonetic, phonemic and environmental. Moreover, [f] and [θ] have about the same low intensity, according to preliminary evidence.[18]

That leaves the formant transitions which, in accord with the earlier discussion of the plosives, may be expected to converge on different loci depending on the place of articulation of the adjacent fricative. Spectrographic measurements confirm that plosive and fricative consonants with the same or proximate places of articulation have nearly the same formant loci (Green, 1958). Complementary research on the identification of synthetic formant transitions with variable loci has not been reported for the fricatives. However, Delattre (1962b) mentions in an address that the voiced counterparts of [f] and [θ] could be contrasted by synthesis without a noise component, varying only the transitions of F2 and F3 of the following vowel. The hand-painted spectrograms which were required are shown in Fig. 48. Also shown are the patterns which synthesized the [z] - [ʒ] contrast in the mid-group of fricatives. Each of the patterns for the four voiced fricatives includes traces of energy, during the frication interval, which lead into the transitions of F2 and F3. The relative importance of these traces, which correspond to friction-formants that are barely visible on the spectrograms for voiced fricatives before vowels, has not been established. Comparison of Figs. 40 and 48 reveals that the F2 transition of labiodental [vɛ] is negative and aimed toward a low frequency locus in the region of that for the bilabials. Similarly, the dental [ðɛ] has a negative transition with

- - - - - - - - - - - - - - - - - - -

Insert Fig. 48 about here

- - - - - - - - - - - - - - - - - - -

slightly higher locus. In the case of [zæ], the fricative has nearly the same point of articulation as the alveolar plosives and hence the same high locus; F2 of [æ] is sufficiently high that its transition is minimal. The palatal [ʒ] apparently required a positive transition as did its neighbors the velar plosives; like them, its locus seems to be in the region of 3,000 cps.

Perhaps the most persuasive evidence that the transitions of the formants of an adjacent vowel and the spectral distribution of the friction both play a role in the identification of a fricative comes from a tape-splicing experiment with natural language. Harris (1958) recorded four English syllables, [fi], [θi], [si], [ʃi], then separated the friction and vocalic segments of each (at a point just before the vowel transitions) and finally combined each of the four types of friction with each of the four vocalic segments. Presumably, the dominant fricative cue would be revealed by a predominance of fricative identifications corresponding to the source of the friction, on the one hand, or to the source of the vocalic portion, on the other. The stimulus array was increased eight-fold by preparing similar bastard syllables for the voiced fricatives and three other vowels, [e, o, u]. When [s] friction was paired with any vocalic portion, the product was identified as [s]. Similarly, [ʃ]friction paired with any set of formant transitions yielded [ʃ]identifications. Corresponding results were obtained with [z] and [ʒ]. The spectral distribution appears to be a more important cue than the formant transitions of an adjacent vowel in identifying the fricatives in the mid-group. In the front group, [f] identifications were predominant whenever the vocalic portion was [f] and only then. A combination of the friction of [f] and the vocalic portion of [θ] evoked mainly [θ] identifications. Similar results were obtained with the voiced fricatives [v] and [ð]. The formant transitions of an adjacent vowel appear to be the more important cue, therefore, in differentiating the fricatives in the front group. If a vocalic portion from the front group is pitted against a friction portion from the mid-group, the product is identified as a fricative from the mid-group. In general, the results support the conclusion that discrimination between fricatives in different groups is based on their friction rather than vocalic portions.

Confirmation of these findings was obtained by synthesis of fricative-vowel combinations with an electrical analog of the vocal tract (Heinz & Stevens, 1961). A schematic representation of a typical acoustic pattern used in the experiment is shown in Fig. 49. The variables were the resonant frequency of the friction and the transitions of the vowel formants. Listener identifications of the synthetic fricatives are shown by a contour plot in Fig. 50. When the fricative portion had an intensity 25 db below that of the vowel, the mid and front groups of fricatives, as well as [ʃ] and [s] within the former, were distinguished by the resonant frequency of the friction whereas [f] and [θ] were distinguished by the formant transitions of the adjacent vowel.

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 49 about here

- - - - - - - - - - - - - - - - - - - -

- -

Insert Fig. 50 about here

- -

The glottal fricatives [h] and [ɦ] are deviant in several respects and require separate consideration. Their inclusion in the class of fricatives is not undisputed (Kurath, 1957) and little has been reported about their acoustic or discriminative properties. A noteworthy exception is a systematic study by Lehiste (1962a) which employed primarily spectrographic techniques. At the outset in examining word-initial [h] , we may note the noise excitation, the presence of energy maxima corresponding to F2 and F3 of the following vowel and the absence of formant transitions (cf. Fig. 45). Based on Lehiste's report, the following additional generalizations may be made. The formant structure of [h] may be said to anticipate that of the following vowel but F1 is typically diffuse and readily observed only if the vowel has a high first formant. The formant frequencies during the fricative segment are usually a little higher than in the following vowel. Finally, the rate of airflow during the production of [h] is about six times as great as that of the following vowel; [h] has the highest sound pressure per unit subglottal pressure of the voiceless fricatives considered here (Strevens, 1960).

Distinctions among the fricatives according to manner of articulation have received very limited study. The systematic increase of vowel duration before voiced plosives and fricatives has been cited in earlier sections and Denes (1955) has demonstrated that this increase is a cue for discrimination between voiced and voiceless fricatives. He also showed that the number of "voiced" identifications increases as the duration of the fricative and its intensity relative to that of the vowel decrease.

The affricates may be distinguished from the fricatives by their plosive component; to a first approximation they may be considered a compound of a plosive and a fricative segment (Jassem, 1962), as one common notation indicates. However, the fricative segment of an affricate is shorter in duration and has a faster rise-time than the corresponding fricative. According to Delattre (1958b) "one perceives unvoiced fricatives when the total noise lasts at least 110 msec, unvoiced affricates when the total noise lasts at least 50 msec and unvoiced plosives when the noise lasts at most 30 msec."

The nasals

Spectrograms of the nasal consonants [m, n] spoken in initial position before [a] by a Russian informant, appear in Fig. 51 along with their amplitude sections, taken at the indicated points during the nasal murmur.[19] The English velar nasal [ŋ], which does not occur in initial position, is not shown. As illustrated by the articulatory diagram in Fig. 52, the nasals combine the lowered velum of nasalized vowels with the oral occlusion of the plosives and, therefore, have some of the

- - - - - - - - - - - - - - - - - - - -
Insert Fig. 51 about here
- - - - - - - - - - - - - - - - - - - -

acoustic properties of both, which were described earlier.

Insert Fig. 52 about here

During the period of "nasal murmur" preceding the formant transitions of the adjacent vowel, the velum is lowered, the oral cavity is sealed and acoustic energy is largely radiated from the nares. The configuration for [m] includes bilabial closure (about 15 cm from the glottis) and a tongue position dependent on the formant frequencies of the following vowel. [n] is characterized by alveolar closure (13 cm from the glottis) and [ŋ] by velar closure (10 cm from the glottis). In discussing the nasalization of vowels earlier, it was pointed out that coupling the nasal cavity to the vocal tract during the production of a vowel in effect added a sidebranch to the oral cavity, which introduced new resonances and antiresonances, increased the damping of formants and reduced their level. Now the situation is reversed: the basic cavity is pharyngo-nasal and the sidebranch is an oral cavity of variable area, depending on the place of articulation of the nasal consonant. For [ŋ], the tongue seals off the oral cavity just a few centimeters away from the nasal port, so the spectrum of [ŋ] may be considered to reflect the resonant properites of the basic cavity (Hattori, Yamamoto & Fujimura, 1958). Detailed spectral measurements with the technique of analysis-by-synthesis (Fujimura, 1962) reveal that the first resonance of this cavity is a broad formant in the region of 300 cps; the extensive damping reflects acoustic losses due to the relatively large surface area and mucosal covering of the nasal passages. Additional formants are observed at approximately 1, 2, and 3 kcs but these have considerably less amplitude (cf. Fig. 51). When the area of the oral cavity is increased for the articulation of [n], an antiresonance occurs in the region of 1,800 cps; when it is increased further for [m], the frequency of the antiresonance is lowered to about 1,000 cps (cf. Fig. 51).

With respect to manner of articulation, then, the nasal murmurs may be distinguished from the vowels by the presence of a single major formant, heavily damped and located at low frequency, and by a marked antiresonance at intermediate frequencies. These detailed spectral properties lead to the gross acoustic difference of a more rapid decline in energy, as a function of frequency, and a lower overall energy associated with the nasals.[20] Evidence for the autonomy of nasality as a cue for manner of articulation comes from the experiment on consonant intelligibility by Miller and Nicely, discussed earlier. When the confusions associated with the nasal consonants, the affricates, and the plosives were sorted separately at each level of the masking noise, nasality had an articulation function similar to that shown for voicing in Fig. 35. Like voicing, information about nasality is transmitted under adverse conditions when other acoustic cues are received poorly or not at all.[21]

With respect to place of articulation, the location, notably, of the antiresonance and of minor resonances permit distinctions among the nasal consonants. Several experiments have provided evidence that these spectral differences can exert discriminative control.[22] When the nasal murmurs were synthesized by electrical analogs of the three vocal tract configurations, each of the consonants was identified with modest consistency by a panel of listeners. Comparable identification distributions were obtained when the nasal murmurs were separated from recordings of English utterances by tape splicing and then presented in isolation to listeners.

The second set of acoustic properties of the nasal consonants, deriving from their kinship with the plosives, consists of the formant transitions of adjacent vowels. Based on articulatory observations, we may expect that transitions associated with [m] before various vowels will converge on the bilabial locus established for [p, b] and, proximately, for the fricatives of the front group; similarly, the formant transitions of [n] are expected to converge on the alveolar locus of [t, d, s, z] and those of [ŋ] on the velar locus of [k, g]. Spectrographic measurements confirm this prediction for the most part (Green, 1958) and experiments with synthetic and altered speech have established that the different formant loci (or, more exactly, their associated transitions) do in fact control discriminations among the nasal consonants. In one of these experiments, hand-painted spectrograms like those shown in Fig. 53 were converted into sound with pattern playback and presented to listeners for identification. The features of the voiced and voiceless plosives have been discussed in an earlier section. The features of the nasals include resonances at 240, 1020, and 2460 cps, which were chosen informally as yielding the best nasal quality but coincide well with spectral measurements of the nasal murmurs in subsequent research. The transition and formant frequency of F2 were the independent variables of the experiment, ranging from +6 to -4 and from 720 to 2760 cps, respectively. The dependent variable was consonant identifications; the distributions are shown in Fig. 54. The identification of a synthetic pattern as [m] occurred most often when the transition of F2 was negative. Reference to Fig. 40 shows that this finding is in accord with the locus hypothesis since the bilabial locus lies below the lowest F2 in the vowel series. Accordingly, the peak of [n] identifications is at zero transition for the second formant of [ɛ], since F2 of this vowel is at the alveolar locus, and it is at positive transitions for the back vowels and negative transitions for the front vowels. Finally, the peak of [ŋ] identifications favors positive transitions since the velar locus is at a higher frequency than the highest F2 of the vowels. The peculiarities of the velar locus before [ɔ, o, u] have been discussed previously.

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 53 about here

- - - - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 54 about here

- - - - - - - - - - - - - - - - - - - -

It is clear from the identification functions in Fig. 54 that the F2 transition of the adjacent vowel is an important cue for the discrimination of nasal consonants. Nakata (1959) came to a similar conclusion with nasal-vowel syllables synthesized on an electrical analog of the vocal tract. Figure 54 also shows, however, that this cue is somewhat less effective in distinguishing among the nasals than among the voiced plosives. Some measure of the generalization among the nasals may be attributable to the use of an identical nasal murmur for all three consonants. It was shown above that the three nasal murmurs have characteristically different spectra and may be discriminated on this basis. The question now arises, as it did with the plosives and fricatives in preceding sections, what are the relative contributions of spectral and transitional cues in discriminating the place of articulation of a nasal consonant? To answer this question, Malecot (1956) interchanged the murmurs and transitions in recorded utterances of [mæ, næ, ŋæ] and [æm, æn, æŋ] by means of tape splicing and presented the various combinations to listeners for identification of the nasal consonants. Table VI summarizes the design and results of the experiment. Pattern 72, for example, consisted of the nasal murmur of [m] recorded in [mæ] plus the formant transitions associated with [ŋ], found in the vowel segment of [ŋæ]. The pattern was identified as [ŋæ] 68 per cent of the time and as [mæ] 16 per cent of the time, indicating greater control of discrimination by the transition cues. In general the data support the conclusion that, in the identification of the nasals according to place of articulation, the formant transitions of an adjacent vowel play a major role and the spectrum of the nasal murmur a secondary but signigicant role.

- - - - - - - - - - - - - - - - - - -
Insert Table VI about here
- - - - - - - - - - - - - - - - - - -

Finally a third type of acoustic cue for the identification of a nasal consonant, effective under special but common conditions, should be noted. In English, syllabic nasal consonants are typically preceded by a homorganic consonant, that is, one with the same place of articulation; thus, "rip'em" [rɪpm̩], "written" [rɪtn̩], "kickin'" [kɪkŋ̍]. Malecot (1960a) recorded the nasal consonants in isolation, and spliced them onto the end of recorded words ending in stops or fricatives with various places of articulation. He found that the nasal syllabic was most often identified by listeners as that nasal consonant which was homorganic with the preceding plosive or fricative, no matter what the actual place of articulation of the nasal segment.

Semivowels and liquids

The semivowels [w, j] and liquids [r, l] may not form as cohesive a group as the plosives, for example, but there are articulatory, distributional and acoustic reasons for considering them together. On articulatory grounds, they differ from the consonants discussed previously by their lesser degree of oral closure. In regard to their distribution in English, they occupy a position adjacent to the vowel nucleus of a syllable (or nucleus plus [r]) and they are the only sounds in English that may intervene between a cluster of two initial consonants and the vowel nucleus of a syllable. The relations among their acoustic and discriminative properties are discussed below.[23] Spectrograms of the semivowels and liquids are presented in Fig. 55. It is apparent that each of the phones has both a steady-state segment and a marked transition. Table VII presents the average formant frequencies of [w, j, l, r] in initial position before a large number of utterances by five American male speakers. The first two formants of [w] have low frequencies and high amplitudes; the third formant is quite weak. This segment of [w] may be likened to a back vowel but F1 and F2 are both lower in frequency than the corresponding formants of any of the English vowels. The formants of [w] are not affected by those of the following vowel but of course the transitions are. The partial similarity between [w] and [u] is matched by

- - - - - - - - - - - - - - - - - - -
Insert Fig. 55 about here
- - - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - - -
Insert Table VII about here
- - - - - - - - - - - - - - - - - - -

the similarity between [j] and [i] in their steady-state segments. [j] has a very low F1 and a very high F2, exceeded only slightly by that of [i]; F3 is relatively weak in amplitude. As in the case of [w], the formants of [j] are not affected by those of the following vowel whereas the transitions span the semivowel and vowel formants. The rise of the transitions of the semivowels is accompanied by a rise in intensity which reaches a sustained maximum at the onset of the vowel steady-state. All three formants of initial [l] are evident in the spectrogram. Their average frequencies in English are given in Table VII. Final [l] differs from initial [l] in that its first two formants are closer together (455 and 795 cps). The formant frequencies do not fluctuate much with changes in the adjacent vowel, although there is some evidence that F2 is higher before front vowels and lower before back vowels. On the other hand, initial [l] tends to increase F1 and F3 of the following vowel, while final [l] is associated with a reduction in F2 of the preceding vowel. Intervocalic [l] has these properties: in words like "freely", where [l] initiates a suffix, it has the acoustic properties of word-initial [l]. In words like "mealy", where [l] precedes the suffix, it has the properties of word-final [l]. In words like "really", both variants of [l] are observed, word-final followed by word-initial. Syllabic [l] ("pickle") resembles word-final [l] in formant structure but it has greater intensity and duration.

The spectrograms of initial [r] show a marked third formant whose frequency is close to that of the second formant; the frequency of F1 is quite low (cf. Fig. 55 and Table VII). This formant structure is not affected by that of the following vowel. Final [r] differs from the initial variant in its higher formant frequencies (455, 1285, 1560) and greater variation in formant structure depending on the preceding vowel. Lehiste (1962b) has also determined the average formant frequencies of [r] and the vowel [ɝ] in the following additional phonetic environments: (3) [ɝ] as syllable nucleus before a final consonant ("burn"); (4) [ɝ] as nucleus without a final consonant ("burr"); (5) the suffix -er (-r) added to vowels ("drawer"); (6)

the suffix -er added to stems ending in r ("bearer"); (7) stem final r before the suffix -y (-ie) ("dearie"); (8) intervocalic r ("arid"); (9) stem-final r before -ing ("boring"); (10) postconsonantal final r (syllabic r) ("author"); (11) postconsonantal r before following vowels ("angry"). All these variants of [r] and the vowel nucleus [ɝ] have common acoustic properties which characterize them as a group and distinguish this group from other speech sounds; the formant pattern of word-final r is representative of the group.

The influence of initial [r] on the following syllable nucleus is minimal. The vowels before final [r] constitute a special case since, in Midwest American English, the 15 stressed syllable nuclei are reduced to 8 in this position. The complex relations between the total vowel system and the subsystem before final [r] are described by Lehiste (1962b).

Experiments with synthetic speech have explored the discriminative function of the following acoustic parameters of the semivowels and liquids: the rate and duration of the transition of the formants, the presence of a third formant, and the starting frequencies of F2 and F3. The first of these parameters, the "tempo" of the formant transitions, has been shown to serve as a cue for manner of articulation, distinguishing the semivowels from the plosives on the one hand and the vowels on the other. The affinity of [w] - [u] and [j] - [i] has already been mentioned. Noting the bilabial and velar places of articulation of [w] and [j], respectively, we may envision a three-term relation in both cases: [b] - [w] - [u] and [g] - [j] - [i]. The identification of synthetic plosives does indeed shift to semivowels and then to vowels when the formant transitions are gradually increased. In order to examine this shift quantitatively, Liberman and his collaborators (1956) prepared hand-painted spectrograms like those shown in Fig. 56. The upper series has the negative F2 transitions appropriate to the bilabial locus,

- - - - - - - - - - - - - - - - - - -
Insert Fig. 56 about here
- - - - - - - - - - - - - - - - - - -

while the lower series has the positive F2 transitions associated with the (high frequency) velar locus. The durations of the transitions in both series were varied from 10 to 300 msec in steps of 10 msec. With the fixed formant frequencies for the vowel [ɛ], a change in the duration of the transitions necessarily involves a change in their rate. When the spectrograms were converted into sound with pattern playback and presented to listeners for identification, the distributions of responses shown in Fig. 57 were obtained. Rapid transitions embedded in the appropriate pattern are consistently identified as plosives. Transitions lasting about 100 msec lead to the identification of semivowels, while very gradual transitions lead to vowel responses.

- - - - - - - - - - - - - - - - - -
Insert Fig. 57 about here
- - - - - - - - - - - - - - - - - -

Since the effects of rate and duration of the transition were confounded in the preceding experiment, a follow-up study was conducted. By expanding the first experiment to include synthesis of five other vowels beside [ɛ], it was possible to vary the rate of transition at each value of duration. The transitions all began at the same frequency and terminated at the variable F2 of the following vowel. For a given duration of transition, then, the front vowels, with their higher F2, required perforce a transition of greater rate than did the back vowels. When these patterns were presented to listeners for identification, the shift in responses from plosives to semivowels occurred systematically, and the distributions for the different vowels were nearly coincident - when plotted against the duration of the transition. When the frequencies of the identifications were plotted as a function of the rate of transition, however, six distinct distributions were obtained, one for each of the vowels. This outcome led the investigators to suggest that the duration of the transitions was the significant cue for discriminations among plosives, semivowels, and vowels-of-changing-color.

A subsequent study, utilizing essentially the same techniques, explored the role of the starting frequencies of the F2 transition and the presence (and slope) of an F3 transition in the identification of the semivowels and liquids. In order to synthesize these speech sounds, the investigators arrived at the patterns shown in Fig. 58 by a long series of informal "cut-and-try" experiments. It is instructive

- - - - - - - - - - - - - - - - - - - -

Insert Fig. 58 about here

- - - - - - - - - - - - - - - - - - - -

to compare these patterns with the spectrograms of natural language shown in Fig. 55 and the discussions of their properties which followed. Pending the more systematic investigations to be described below, O'Connor and his coworkers (1957) chose to include a third formant for the liquids and not for the semivowels. This accords with their spectral distributions in natural language. Similar agreement obtains with regard to the presence of steady-state segments preceding the transitions. These segments are located at the consonant loci for F2: low frequency for the bilabials, intermediate frequency for the alveolars, and high frequency for the velars - in this case [w], [r, l] and [j], respectively (cf. Table VII). It is not surprising, then, that [r] and [l], with their similar F2 loci, require a distinctive F3 for their discrimination. It was found that the optimal onset frequency for F3 was close to that for F2 in the case of [r], but for [l] the starting frequencies of F2 and F3 were best separated widely. This concurs with their spectral patterns in natural language (cf. Table VII). With respect to the loci of the first formant onsets, the best synthesis was obtained when F1 of the semivowels and liquids was located between 240 and 360 cps; this accords with the measurements of F1 in natural language, presented in Table VII. The parameter of transition duration has been discussed above, where durations in the order of 100 msec were found to be optimal for the semivowels. The present exploratory work led to the adoption of the same value for the transitions of the liquids.

Having arried at these basic patterns for the synthesis of the semivowels and liquids, the experimenters varied their parameters, as shown in Fig. 59, and presented the array of synthetic speech sounds to listeners for identification. The

- - - - - - - - - - - - - - - - - - -
Insert Fig. 59 about here
- - - - - - - - - - - - - - - - - - -

results are shown by the family of graphs in Fig. 60; the columns correspond to the predominant identification of the initial speech sound, the rows separate the

- - - - - - - - - - - - - - - - - - -
Insert Fig. 60 about here
- - - - - - - - - - - - - - - - - - -

patterns preceding each of the three vowels. Each graph shows the per cent of the indicated identifications that occurred at each of the onset frequencies of F2; the parameter is the presence and slope of F3. Inspection of Fig. 60 shows, first, that the onset frequency of the F2 transition distinguishes [w], [r, l] and [j]. The maximum number of identifications of [wa], for example, occurred at the lowest onset frequencies, that for [ra] and [la] at intermediate values, and the maximum for [ja] occurred at the highest onset frequency employed. This result is in accord with the exploratory work described earlier, with the spectrographic measurements already presented, and with the fundamental concept of formant locus. The response distributions in Fig. 60 also show that the presence of a third formant does not contribute to the identification of [w] or [j] (in agreement with the weak F3 found on their spectrograms) whereas a rising F3 considerably enhances the identification of [r] and a straight F3 improves the identification of [l].

The results of a similar experiment, concerned with the discriminative properties of [w, j, r, l] in intervocalic position, are summarized by the identification contours in Fig. 61. The variables are the onset frequencies of the second and third formants; each of the shaded areas encompasses that range of onset frequencies for which the indicated identification was given by more than half of the listeners. Confirming several earlier findings, it will be observed that the F2 locus for the bilabial [w] is at low frequency, that for alveolar [r, l] at intermediate frequency,

Insert Fig. 61 about here

while the optimal locus for the velar [j] (=[y]) is at high frequency. Also in support of earlier results, the onset frequency of F3 has little effect on the identification of [w, j] whereas a wide separation of F2 and F3 onsets is clearly favored for [l] and a small separation for [r].

We may conclude that of the various acoustic properties of the semivowels and liquids their primary "manner" cue, controlling discriminations between them as a group and other speech sounds, is the duration of their formant transitions, while their primary "place" cues, controlling discriminations within the group, are the loci of their second and third formants.

Prosodic features

Prosodic or suprasegmental features are characterized by acoustic invariances that extend over several segments in the stream of speech. Four overlapping phonetic problems are encompassed under this heading: intonation, tempo, juncture and stress.

The phonetic and linguistic literature is replete with conflicting assertions about the structure and function of intonation but there are few objective and systematic investigations which can be reported here.[24] The first firm step toward the description of acoustic invariances in intonation seems to have been taken by Delattre (1961, 1962c, 1962d) with spectrographic measurements of English, Spanish, French and German conversations. Delattre (1962c) points out that these measurements

> "generally confirm the subjective notions on which [several] linguists agree, but they provide in addition
>
> a. the statistical element which indicates the amount of leeway that exists within an acceptable pronumciation, and

b. the detailed shape of actual contours which is needed to complement the schematic or numerical notations that are commonly used" (p. 234).

Representative contours for declarative intonation in four languages are shown in Fig. 62. As the Figure indicates, spectrographic tracings of the fundamental frequency during conversation seemed to cluster in three groups. Spanish and English, for example, both show falling intonation over the final segments of a declarative sentence but they differ otherwise, with ascent characterizing Spanish and descent American English. Of the 139 "continuation groups" measured for the Spanish speaker, only 25 showed a rise that was shorter than or equal to the following descent; in 207 measurements for an English speaker only 22 showed a rising intonation contour.[25]

- - - - - - - - - - - - - - - - - - -
Insert Fig. 62 about here
- - - - - - - - - - - - - - - - - - -

As in the case of intonation, only a start has been made in describing the acoustic invariances associated with the tempo of speech and their effects on the discriminative behavior of listeners. The gross distributional properties of speech duration have received some study; it is known, for example, that changes in the rate of "connected" discourse may be attributed largely to the pauses therein, since the rate of speaking in normal conversation, pauses aside, varies only slightly, some four to six syllables per second (Goldman-Eisler, 1956. Also see Fonagy & Magdics, 1960). The rate of speaking has been employed as a dependent variable in several studies (especially Goldman-Eisler, 1958a and Goldiamond, 1962)--but, again, these results are not pertinent to the present question: are there characteristic temporal patterns of energy distribution associated with utterances in a language and, if so, do these patterns differ systematically within and between languages?

Lounsbury (1954) hypothesized that there are close relations between hesitation pauses in speech and the transitional probabilities or constituent boundaries at which these pauses occurred. These relations have been demonstrated indirectly by several studies. In one experiment (Goldman-Eisler, 1958b) listeners were asked to fill in the gaps in recorded sentences from which words with high or with low transition probabilities had been deleted. It took listeners 20 secs, on the average, to come up with a word that was well-determined contextually and 37 secs to give one that was poorly determined.[26] Some of the words that had been deleted were uttered fluently by the original speakers while others followed a hesitation pause. When the pauses of speakers and listeners were compared, it was found that replacing a word that had been preceded by a pause took significantly longer than replacing one that had been uttered fluently in the stream of speech.

There is some evidence, then, that variations in tempo are not entirely idiosyncratic and that invariances may be found associated with the grammatical structure of the utterances in a language. On the other hand, it seems likely that tempo, perhaps more than intonation and certainly more than the segmental features of speech, does not comprise a limited set of relatively constrained patterns. Instead, tempo may be largely under the control of nonliguistic variables. Goldman-Eisler (1961b) has described some of the differences in tempo in interviews and spontaneous and learned descriptions and summaries. A description of the degree to which the temporal patterning of speech is conventionalized by the language community requires parcelling out the effects of physiological, linguistic, and environmental variables through controlled experiments with numerous speakers, utterances and languages. This has not been accomplished as yet.

What is known objectively about the problem of juncture may largely be attributed to a recent study by Lehiste (1960). We begin with the concept of a bounded sequence of speech segments. It has been pointed out recurrently in the preceding pages that the acoustic properties of a speech sound may differ systematically when it occurs in word-initial, medial, and word-final positions. In English, initial variants are typically longer than their medial counterparts. Initial voiceless plosives are aspirated but their medial and final counterparts are not. Initial

[w] is voiced but following a plosive it is not. Initial [l] has a different formant pattern than final [l]. And so forth. A bounded sequence, then, begins with the word-initial variant of some speech sound, which is followed by a sequence of medial segments, and terminates with the word-final variant of a speech sound. A bounded sequence may also be characterized by a rapid rise in intensity followed by a gradual decay.

When a set of segments which constitute a bounded sequence are spoken so as to form two or more bounded sequences, we say that (internal open) juncture has been introduced. In this case, two of the medial variants in the original sequence are altered; one becomes a word-final variant, the other a word-initial variant. Consider "tulips" vs. "two lips" as an example. Spectrographic measurements with two utterances from each of three speakers showed that in the single bounded sequence, the aspiration of [t] lasted about 50 msec, the [u] 100 msec, the [l] 50 msec. When medial [ul] became word-final and word-initial, respectively, their durations changed to 150 and 80 msec, respectively. Other examples of English juncture studied by Lehiste include "a nice man - an iceman", "it sprays - it's praise", "the sun's rays meet - the sons raise meat", "white shoes - why choose", etc. In each case, the medial variants of speech sounds change to their word-initial or word-final variants by slight alterations in their duration, intensity, formant pattern, and the like.[27]

The acoustic and discriminative differences between utterances such as "Don't rebel" and "Don't be a rebel", variously called differences in stress, accent, and prominence, have received more extensive study than the other prosodic features. Fundamental frequency, intensity and duration were identified as acoustic correlates of word stress in an experiment in which twenty five stress contrasts like that just cited were recorded and analyzed electronically (Lieberman, 1960). Table VIII summarizes the results of these measurements. Of the 50 bisyllabic English words recorded by speaker no. 1, for example, the stressed syllable exceeded the following unstressed syllable in frequency, intensity and duration 48, 43 and 35 times, re-

Insert Table VIII about here

spectively, and its unstressed counterpart in the other word of the stress pair 43, 45, and 40 times, respectively. In general the stressed syllable had a higher fundamental frequency than the unstressed syllable of the same utterance in 90 per cent of the cases, a higher peak envelope amplitude in 87 per cent and a longer duration in 66 per cent. The corresponding percentages when the stressed syllable is compared with its unstressed counterpart are 72, 90 and 70. It is noteworthy that in those cases in which the higher fundamental frequency fell on the unstressed syllable there was a proportionate increase in the amplitude of the stressed syllable, and conversely. In other words, a "trading" relation was observed between two of the cues that contribute to the discrimination of stress. More often, however, a greater fundamental frequency was the reliable accompaniment of stress.

A fourth acoustic difference between the syllables of words such as "rebel" is called vowel reduction. The vowels which occur in the unstressed syllable have a more central location in the acoustic vowel diagram; their formant frequencies, especially F2, "undershoot" the more extreme and differentiated values observed in isolated stressed utterances. Since the vowel schwa [ə] has a relatively central position in the acoustic vowel diagram, reduced or unstressed vowels are sometimes said to have a schwa-like quality. We have seen earlier that vowels are also reduced from their values in isolated stressed utterances when they occur in the stream of speech (Fig. 14). Lindblom (1963) has linked the two observations of vowel reduction with the duration of the vowels involved. His argument is that each vowel has an invariant set of target formant-frequencies (cf. Fig. 11). The extent to which the articulators reach their target positions depends on the number of targets they must approximate in a given period of time. When the vowel duration is long, the targets may be reached and even maintained. When the vowel duration is short, as in the unstressed syllable of an isolated word or, more generally, in the rapid stream of speech, the articulators undershoot the target and move quickly on to the next.

Lindblom demonstrated the comparable relations between vowel reduction and vowel duration under the two conditions, in unstressed syllables and in rapid speech, in this way. In one experiment he placed a test syllable, containing one of eight vowels, in various positions in Swedish sentences, so that the syllable received varying degrees of stress. We have seen previously that changes in the stress of a syllable are accompanied by changes in its duration and Fry (1955) has shown that these changes are largely attributable to changes in the duration of the vowel nucleus of the syllable. Therefore, the formant frequencies of each vowel could be measured at several values of its duration. The results for each of the eight vowels in one of the syllables are shown in Fig. 63. It will be observed that with decreasing vowel duration, the formant frequencies increasingly "undershoot" their target values (dashed line). The degree of undershoot or reduction is a function of the extremity of the target, its departure from the "neutral" F-pattern of [ə] . [i], with a very high F2, and [ʊ], with a very low F2, show the greatest reduction; at short durations, these vowels move a considerable distance toward the center of the acoustic vowel diagram. The other vowels show varying degrees of reduction, depending on the extremity of their target formant-frequencies.

- - - - - - - - - - - - - - - - - - -

Insert Fig. 63 about here

- - - - - - - - - - - - - - - - - - -

As a further demonstration that vowel reduction in stress may be attributed to the reduced time allotted the articulators to reach the vowel targets, Lindblom performed a second experiment in which a speaker repeatedly produced the test syllables at varying rates, synchronous with the clicks of a pulse generator. The formant frequencies of one of the vowels, determined at several values of its duration, are shown by the circles in Fig. 64. The increase in the relatively low

F2 of this vowel, observed at short durations, constitutes vowel reduction. The

- - - - - - - - - - - - - - - - - - -
Insert Fig. 64 about here
- - - - - - - - - - - - - - - - - - -

curves shown in Fig. 64 are redrawn from Fig. 63. They are the functions obtained for this syllable when duration and vowel reduction covaried with varying degrees of stress. The good agreement between the two sets of data supports the assumption that vowel reduction is related directly to vowel duration and only indirectly to stress, and it supports the interpretation which attributes the direct relation to temporal factors in articulation.

With four acoustic differences between stressed and unstressed syllables identified, we may now inquire into their discriminative properties. In order to assess the contribution of each of the acoustic variables to the discrimination of stress, each may be varied independently in the preparation of synthetic speech. With this technique, Fry (1955) was able to conclude that duration was a more significant cue than intensity in discriminating between two English words with contrastive stress. Using hand-painted spectrograms, he synthesized stress pairs such as obJECT and OBject on pattern playback, by systematically varying the ratios of intensity and duration between the two syllables of each word. Figure 65 shows that, as the duration and intensity ratios were increased, a group of listeners

- - - - - - - - - - - - - - - - - - -
Insert Fig. 65 about here
- - - - - - - - - - - - - - - - - - -

reported a corresponding increase in the number of words with stress on the first syllable. Since Fry employed duration and intensity ranges as wide as any he found in measurements of stress contrast in natural language, and since these ranges effected a 70 to 29 percent increase in stress judgments respectively, he concluded that duration was the more important cue for the discrimination of English stress.[28]

Several investigators have gone beyond Fry's conclusion and have argued that heightened intensity may be an acoustic concomitant of stress but it is not a discriminative cue for stress judgments in English. Mol and Uhlenbeck (1956) point out that if the normal frequency and duration cues are available in a tape-recorded, stressed disyllable, either syllable may be reproduced at any relative intensity that is audible without affecting the judgment of which syllable is stressed. Bolinger (1958a) synthesized several samples of an English test sentence so that the stressed syllables had an appropriate level of fundamental frequency and a variable level of intensity. He then presented pairs of these sentences to listeners who were asked to say which of the two sentences in each pair sounded better. Raising the intensity of the pitch-accented syllable did not improve its quality, according to judgments by the listeners, and more than a small increase in intensity yielded a reduction in quality.[29]

The preeminent importance of fundamental frequency as a cue for stress discriminations has been demonstrated in several experiments. Garding and Gerstman (1960) synthesized the sentence "Where's he living now?" 17 times, each time placing the intonation peak at a different point in the sentence, as illustrated schematically at the top of Fig. 66. Listeners were asked which of the syllables: Where's-he-liv-ing-now, was stressed in each of the recordings of the sentence. Their responses are tabulated at the bottom of Fig. 66, where the number of votes for each syllable (rows) is listed as a function of the location of the intonation peak (columns). It will be observed that judgments are nearly unanimous for "where's", "liv-" and "now" within steps 1-4, 7-9, and 13-17, respectively. These findings demonstrate the control of stress judgments in English exerted by an increase in fundamental frequency.

- - - - - - - - - - - - - - - - - -
Insert Fig. 66 about here
- - - - - - - - - - - - - - - - - -

The results of the preceding experiment also indicate that the precise location of an intonation peak within a syllable is not critical for its identification as stressed. Further examination of the data in Fig. 66 shows that when the intonation peak is equidistant between two syllables, it is usually the earlier syllable that receives the stress judgments. This suggests that a rise in pitch may be a more effective cue for stress judgments in English than a fall in pitch, when the same pitch maximum is attained in either case. Bolinger was led to a similar conclusion by his experiments with synthetic speech, but he emphasizes that "while the upward obtrusion is basic, pitch prominence need not be merely upward, as commonly supposed, but may take other directions" (1958b, p. 149). The effects of the form of the intonation contour on the discrimination of stress were studied by Fry (1958) with synthetic bisyllabic words to which the contours schematized in Fig. 67 were applied. With pattern A, for example, the fundamental frequency fell linearly

- - - - - - - - - - - - - - - - - -
Insert Fig. 67 about here
- - - - - - - - - - - - - - - - - -

from 130 cps to 97 cps during the first syllable and remained at the frequency baseline during the second syllable. The words synthesized with this contour were identified as nouns (PERmit, CONtract, etc), that is, they were reported to be stressed on their first syllable, in 80 per cent of the listeners' judgments. When the contour applied to the bisyllable was reversed, as in pattern B, then in complementary fashion, the second syllable was judged as stressed 76 per cent of the time. Rows D, E and F of Fig. 67 show that a syllable with rising intonation is generally identified as stressed no matter what the fundamental frequency of the adjacent syllable (within the range studied). In general 61 per cent of rising syllables and 64 percent of falling syllables were judged stressed. Of all the syllables with a linear frequency change, 62 per cent were judged stressed; of those with a curvilinear change, 72 per cent were reported stressed. The direction of change and its linearity do not emerge, then, as major parameters of the effect of intonation on judgments of stress. What is revealed as a major factor is change itself: of those patterns that contained both a level and an inflected syllable, 33 per cent of the level syllables and 66 per cent of the inflected syllables were reported stressed. In another in this series of experiments, not represented in Fig. 67, Fry found that if both syllables have a level intonation, the syllable with the higher fundamental frequency is judged as stressed.

We have seen previously that the experimental evidence relegates intensity to a position of less importance in the determination of judgments of stress. Research findings on vowel reduction are insufficient to establish its position in the hierarchy of cues. Of the remaining competitors, fundamental frequency emerges as the predominant acoustic variable. In an experiment by Rigault (1962), for example, the word "Papa" was synthesized eight times, with the second syllable constant and the first assuming one of two values on each of three parameters: fundamental frequency, 110 or 210 cps; intensity, 0 or -20 db (relative); duration, 200 or 80 msec. (The frequency of the final syllable fell from 150 to 110 cps, the intensity was -10 db and the duration 200 msec). Forty French listeners identified which of the two syllables was stressed in each of the synthetic patterns; Table IX shows their judgments as a function of the parameter values of the first syllable.

- -

Insert Table IX about here

- -

The four words with a high fundamental frequency in the first syllable came in first in the rank ordering by a large margin. Table X displays the difference in the per cent of stress judgments for all pair-comparisons of the eight words,

- -

Insert Table X about here

- -

facilitating an assessment of the relative importance of the three variables. It will be observed, first of all, that the greatest reductions in the identification of stress are associated foremost with a reduction in the fundamental frequency. A more detailed examination of the data shows that when one variable is changed, it is the lowering of frequency which leads to the greatest diminution in the reported stress of the first syllable (nos. 3, 8, 10, 14). When two parameters of the first syllable are changed, the eight cases where frequency is lowered yield the greatest decrement in stress identifications (nos. 1, 2, 6, 9, 11, 12, 13, 15). An increase in duration (11, 13) or intensity (12, 15) did not counterbalance the reduction in frequency with the parameter values employed in this experiment. When three parameters are changed, the reduction in stress judgments is most marked in those cases (4, 5, 7) in which frequency was lowered. An increase in both intensity and duration (16) offsets the reduction in stress caused by a reduction in fundamental frequency. Comparable results were obtained by Rigault in a subsequent experiment in which the test word was embedded in a synthesized sentence. From these findings we may conclude that of the four acoustic correlates of stressed syllables, it is a prominence in fundamental frequency that is the dominant cue in controlling discrimination of stress.[30]

Epilogue

In a discussion of the discrimination of stress in English, Wang has written:

> "It is implausible to suggest that the listener's judgments depend entirely on acoustical cues for their evaluation of English stress in the light of the complexity and elusiveness of these cues. The conclusion must be that he relies heavily on some set of abstract predictive rules, couched in an overall grammar, which guides him in his interpretation of stress and which can allow for marked deviations in the acoustical signal" (1962, p. 76).

The same conclusion must be acknowledged concerning the discrimination of the segmental features of speech. Throughout this survey we have emphasized the interaction between the acoustic properties of adjacent segments but we have omitted any consideration of the fact, which follows from this interaction, that the cues for the discrimination of a given segment lie largely within other adjacent segments. More generally, we have set aside a discussion of the redundancy of cues in the stream of speech, a redundancy that derives from the severe constraints on the concurrent features and temporal patterns of segments that may occur in a given language. There is ample reason to believe that the sources of this redundancy or, in other words, predictabilty of the segmental and suprasegmental features of speech are largely responsible for discriminations among utterances in natural language. These sources of predictability are codified by the phonological and syntactic rules of grammar.

References

Abramson, A. S. The vowels and tones of standard Thai: acoustical measurements and experiments. *Int. J. Amer. Ling.*, 1962, 28, No. 2, Part III.

Andresen, B. S. On the perception of unreleased voiceless plosives in English. *Lang. & Speech*, 1960, 3, 109-119.

Bascom, . Tonomechanics of Northern Tepuhan. *Phonetica*, 1959, 4, 77-88.

Beresford, R. An outline of research information on the synthesis of nasal consonants. Unpublished report, 7/Gen/1421, Phonetics Dept., Edinburgh Univer., June, 1963.

Bolinger, D. L. On intensity as a qualitative improvement of pitch accent. *Lingua*, 1958, 7, 175-182. (a)

Bolinger, D. L. A theory of pitch accent in English. *Word*, 1958, 14, 109-149. (b)

Bolinger, D. L., & Gerstman, L. J. Disjuncture as a cue to constructs. *Word*, 1957, 13, 246-255.

Carterette, E. C., & Møller, A. The perception of real and synthetic vowels after very sharp filtering. Unpublished report, STL-QPSR-1/962, Royal Inst. Tech., Stockholm, March, 1962. Pp. 30-40.

Chomsky, N. Review of "Fundamentals of Language". *Int. J. Amer. Ling.*, 1957, 23, 234-242.

Cooper, F. S. Some instrumental aids to research on speech. In *Report 4th ann. round table meeting ling. & lang. teach*. Washington, D. C.: Inst. Lang. Ling., Georgetown Univer., 1953. Pp. 46-53.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.*, 1952, 24, 597-606.

Delattre, P. C. Les attributs acoustiques de la nasalité vocalique et consonantique. *Stud. Ling.*, 1954, 8, 103-109.

Delattre, P. C. Unreleased velar plosives after back-rounded vowels. _J. acoust. Soc. Amer._, 1958, 30, 581-582 (a)

Delattre, P. C. Les indices acoustique de la parole: Premier rapport. _Phonetica_, 1958, 2, 108-118, 226-251. (b)

Delattre, P. C. La leçon d'intonation de Simone de Beauvoir, étude d'intonation déclarative comparée. _French Rev._, 1961, 35, 59-67.

Delattre, P. C. Some factors of vowel duration and their cross-linguistic validity. _J. acoust. Soc. Amer._, 1962, 34, 1141-1143. (a)

Delattre, P. C. Le jeu des transitions des formants et la perception des consonnes. In A. Sovijarvi & P. Aalto (Eds.), _Proced. 4th Int. Cong. Phon. Sci._ The Hague: Mouton, 1962. Pp. 407-417. (b)

Delattre, P. C. A comparative study of declarative intonation in American English and Spanish. _Hispania_, 1962, 45, 233-241. (c)

Delattre, P. C. _The general phonetic characteristics of languages_. Boulder, Colo.: Author, 1962. (d)

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. _J. acoust. Soc. Amer._, 1955, 27, 769-773.

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. An experimental study of the acoustic determinants of vowel color: observations on the one- and two-formant vowels synthesized from spectrographic patterns. _Word_, 1952, 8, 195-210.

Denes, P. Effect of duration on the perception of voicing. _J. acoust. Soc. Amer._, 1955, 27, 761-764.

Dickson, D. R. An acoustic study of nasality. _J. Speech Hear. Res._, 1962, 5, 103-111.

Dreher, J. J. A comparison of native and acquired language intonation. Unpublished doctoral dissertation, Univer. of Michigan, 1951.

Durand, Marguerite. La perception des consonnes occlusives. _Stud. Ling._, 1954, 8, 110-123.

Durand, Marguerite. De la perception des consonnes occlusives: questions de la sonorité. Word, 1956, 12, 15-34.

Dunn, H. K. Methods of measuring vowel formant bandwidths. J. acoust. Soc. Amer., 1961, 33, 1737-1746.

Fairbanks, G., & Grubb, Patti. The psychophysical investigation of vowel formants. J. Speech Hear. Res., 1961, 4, 203-219.

Fairbanks, G., House, A. S., & Stevens, E. L. An experimental study of vowel intensities. J. acoust. Soc. Amer., 1950, 22, 457-459.

Fant, G. Acoustic theory of speech production. The Hague: Mouton, 1960.

Fischer-Jørgensen, Eli. Acoustic analysis of stop consonants. Misc. Phon., 1954, 2, 42-59.

Fletcher, H. Speech and hearing in communication. New York: Van Nostrand, 1953.

Fonagy, I., & Magdics, K. Speed of utterance in phrases of different length. Lang. & Speech, 1960, 3, 179-192.

Fry, D. B. Duration and intensity as physical correlates of linguistic stress. J. acoust. Soc. Amer., 1955, 27, 765-768.

Fry, D. B. Experiments in the perception of stress. Lang. & Speech, 1958, 1, 126-152.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. Lang. & Speech, 1962, 5, 171-189.

Fujimura, O. Analysis of nasal consonants. J. acoust. Soc. Amer., 1962, 34, 1865-1875.

Garding, Eva., & Gerstman, L. J. The effect of changes in the location of an intonation peak on sentence stress. Stud. Ling. 1960, 14, 57-59.

Goldiamond, I. The maintenance of ongoing fluent verbal behavior and stuttering. J. Mathetics, 1962, 1, 57-95.

Goldman-Eisler, Frieda. The determinants of the rate of speech output and their mutual relations. J. psychosom. Res., 1956, 1, 137-

Goldman-Eisler, Frieda. Speech analysis and mental processes. Lang. & Speech, 1958, 1, 59-75. (a)

Goldman-Eisler, Frieda. Predictability of words in context and the length of pauses in speech. Lang. & Speech, 1958, 1, 226-231. (b)

Goldman-Eisler, Frieda. Hesitation and information in speech. In Proced. 4th London sympos. on info. theory. New York: Academic Press, 1961. (a)

Goldman-Eisler, Frieda. Continuity of speech utterance, its determinants and its significance. Lang. & Speech, 1961, 4, 220-231. (b)

Goodman, A. C. Imitation of intonation patterns. Unpublished doctoral dissertation, Univer. of Michigan, 1952.

Green, P. S. Consonant-vowel transitions: a spectrographic study. Stud. Ling., 1958, 12, 57-105.

Hadding-Koch, K. Acoustico-phonetic studies in the intonation of southern Swedish. Travaux L'Inst. Phonét. Lund, 1961, No. 3.

Halle, M. The strategy of phonemics. Word, 1954, 10, 77-89.

Halle, M. Phonology in generative grammar. Word, 1962, 18, 54-72.

Halle, M., Hughes, G. W., & Radley, J. P. Acoustic properties of stop consonants. J. acoust. Soc. Amer., 1957, 29, 107-116.

Harris, Katherine S. Cues for the discrimination of American English fricatives in spoken syllables. Lang. & Speech, 1958, 1, 1-7.

Harris, Katherine S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Effect of third formant transitions on the perception of the voiced stop consonants. J. acoust. Soc. Amer., 1958, 30, 122-126.

Harris, Z. S. Methods in structural linguistics. Chicago: Univer. Chicago Press, 1951.

Hattori, S., Yamamoto, K., & Fujimura, O. Nasalization of vowels in relation to nasals. J. acoust. Soc. Amer., 1958, 30, 267-274.

Hecker, M. H. L. Studies of nasal consonants with an articulatory speech synthesizer. J. acoust. Soc. Amer., 1962, 34, 179-188.

Heffner, R-M. S. General phonetics. Madison: Univer. Wisconsin Press, 1952.

Heinz, J. M., & Stevens, K. N. On the properties of voiceless fricative consonants. J. acoust. Soc. Amer., 1961, 33, 589-596.

Hoffman, H. S. Study of some cues in the perception of the voiced stop consonants. J. acoust. Soc. Amer., 1958, 30, 1035-1041.

Hogan, D. L. Some studies of fricatives using automated analysis by synthesis techniques. In C. G. M. Fant (Ed.), Proc. Speech Commun. Sem., Speech Trans. Lab. Vol. 1. Stockholm: Royal Inst. Tech., 1963.

Holbrook, A., & Fairbanks, G. Diphthong formants and their movements. J. Speech Hear. Res., 1962, 5, 38-58.

House, A. S. Analog studies of nasal consonants. J. Speech Hear. Disord., 1957, 22, 190-204.

House, A. S. Formant bandwidth and vowel preference. J. Speech Hear. Res., 1960, 3, 3-8.

House, A. S. On vowel duration in English. J. acoust. Soc. Amer., 1961, 33, 1174-1178.

House, A. S., & Fairbanks, G. The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. acoust. Soc. Amer., 1953, 25, 105-113.

House, A. S., Paul, A. P., Stevens, K. N., & Arnold, Jane, B. Acoustical description of syllabic nuclei: data derived by automatic analysis procedures. In C. G. M. Fant (Ed.), Proc. Speech Commun. Sem., Speech Trans. Lab. Vol. 1. Stockholm: Royal Inst. Tech., 1963.

House, A. S., & Stevens, K. N. Analog studies of the nasalization of vowels. J. Speech Hear. Disord., 1956, 21, 218-232.

House, A. S., & Stevens, K. N. Estimation of formant bandwidths from measurements of transient response of the vocal tract. J. Speech Hear. Res., 1958, 1, 309-315.

Householder, F. W. Unreleased /ptk/ in American English. In M. Halle (Ed.), *For Roman Jakobson*. The Hague: Mouton, 1956. Pp. 235-244.

Hughes, G. W., & Halle, M. Spectral properties of fricative consonants. *J. acoust. Soc. Amer.*, 1956, 28, 303-310.

Jakobson, R., Fant, C. G. M., & Halle, M. *Preliminaries to speech analysis*: *the distinctive features and their correlates*. Cambridge, Mass.: Acoustics Lab., Mass. Inst. Tech., 1952.

Jassem, W. The phonology of Polish stress. *Word*, 1959, 15, 269-

Jassem, W. The acoustics of consonants. In A. Sovijarvi & P. Aalto (Eds.), *Proced. 4th Int. Cong. Phon. Sci.* The Hague: Mouton, 1962. Pp. 50-72.

Jassem, W. Noise spectra of Swedish, English, and Polish fricatives. In C. G. M. Fant (Ed.), *Proc. Speech Commun. Sem.*, *Speech Trans. Lab.* Vol. 1. Stockholm: Royal Inst. Tech., 1963.

Joos, M. Acoustic phonetics. *Language*, 1948, 24 (suppl.), 1-136.

Koenig, W. A new frequency scale for acoustic measurements. *Bell Lab. Rec.*, 1949, 27, 299-301.

Kurath, H. The binary interpretation of English vowels: a critique. *Language*, 1957, 33, 111-122.

Ladefoged, P. *Elements of acoustic phonetics*. Chicago: Univer. Chicago Press, 1962.

Lane, H. L. The motor theory of speech perception: a critical review. *Psychol. Rev.*, in press.

Lehiste, Ilse. An acoustic-phonetic study of internal open juncture. *Phonetica*, 1960, 5 (suppl.).

Lehiste, Ilse. Review of P. S. Green, "Consonant-vowel transitions: a spectrographic study". *Phonetica*, 1961, 6, 248-250. (a)

Lehiste, Ilse. Some acoustic correlates of accent in Serbo-Croatian. *Phonetica*, 1961, 7, 114-147. (b)

Lehiste, Ilse. Study of [h] and whispered speech. *J. acoust. Soc. Amer.*, 1962, 34,

742. (a)

Lehiste, Ilse. Acoustical characteristics of selected English consonants. Unpublished report, AD-282765, Univer. Michigan, July, 1962. (b)

Lehiste, Ilse., & Peterson, G. E. The identification of filtered vowels. Phonetica, 1959, 4, 161-177. (a)

Lehiste, Ilse., & Peterson, G. E. Vowel amplitude and phonemic stress in American English. J. acoust. Soc. Amer., 1959, 31, 428-435. (b)

Lehiste, Ilse., & Peterson, G. E. Some basic considerations in the analysis of intonation. J. acoust. Soc. Amer., 1961, 33, 419-425. (a)

Lehiste, Ilse., & Peterson, G. F. Transitions, glides and diphthongs. J. acoust. Soc. Amer., 1961, 33, 268-277. (b)

Liberman, A. M. Some results of research on speech perception. J. acoust. Soc. Amer., 1957, 29, 117-123.

Liberman, A. M., Delattre, P., & Cooper, F. S. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. Amer. J. Psychol., 1952, 65, 497-516.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. Lang. & Speech, 1958, 1, 153-167.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol. Monogr., 1954, 68, No. 8 (Whole No. 379).

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. exp. Psychol., 1956, 52, 127-137.

Licklider, J. C. R., & Miller, G. A. The perception of speech. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951. Pp. 1040-1074.

Lieberman, P. Some acoustic correlates of word stress in English. J. acoust. Soc.

Amer., 1960, 32, 451-454.

Lindblom, B. Spectrographic study of vowel reduction. *J. acoust. Soc. Amer.*, 1963, 35, 1773-1781.

Lisker, L. The distinction between /æ/ and /ɛ/: a problem in acoustic analysis. *Language*, 1948, 24, 397-407.

Lisker, L. Minimal cues for separating /w, r, l, y/ in intervocalic position. *Word*, 1957, 13, 256-267. (a)

Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 1957, 33, 42-49. (b)

Lotz, J., Abramson, A. S., Gerstman, L., Ingemann, Frances., & Nemser, W. J. The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: a tape cutting experiment. *Lang. & Speech*, 1960, 3, 71-77.

Lounsbury, F. G. Pausal, juncture, and hesitation phenomena. In C. E. Osgood & T. A. Sebeok (Eds.), Psycholinguistics: a survey of theory and research problems. *J. abnorm. soc. Psychol.*, 1954, 49 (suppl.), 98-101.

Luce, R. D., Bush, R. R., & Galanter, E. *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.

Maclay, H., & Osgood, C. E. Hesitation phenomena in spontaneous English speech. *Word*, 1959, 15, 19

Magdics, K. Research on intonation during the past 10 years. *Acta Linguistica*, 1963, 13, 133-165.

Malécot, A. Acoustic cues for nasal consonants: an experimental study involving a tape-splicing technique. *Language*, 1956, 32, 274-284.

Malécot, A. The role of releases in the identification of released final stops: a series of tape cutting experiments. *Language*, 1958, 34, 370-380.

Malécot, A. Nasal syllabics in American English. *J. Speech Hear. Res.*, 1960, 3, 268-274. (a)

Malécot, A. Vowel nasality as a distinctive feature in American English. *Language*, 1960, 36, 222-229. (b)

Malmberg, B. Observations on Swedish word accent. *Stud*. *Ling*. 1956, 10, 1-44.

Mártony, J. C., Cederlund, C., Liljencrants, J., & Lindblom, B. On the analysis and synthesis of vowels and fricatives. In A. Sovijarvi & P. Aalto (Eds.), *Proced*. *4th* *Int*. *Cong*. *Phon*. *Sci*. The Hague: Mouton, 1962. Pp. 208-213.

Miller, G. A. The perception of speech. In M. Halle (Ed.), *For* *Roman* *Jakobson*. The Hague: Mouton, 1956. Pp. 353-359.

Miller, G. A., & Nicely, Patrica E. An analysis of perceptual confusions among some English consonants. *J*. *acoust*. *Soc*. *Amer*., 1955, 27, 338-352.

Miller, R. L. Auditory tests with synthetic vowels. *J*. *acoust*. *Soc*. *Amer*., 1953, 25, 114-121.

Mol, H., & Uhlenbeck, E. M. The linguistic relevance of intensity in stress. *Lingua*, 1956, 5, 205-213.

Mostofsky, D. Conference on stimulus generalization. Unpublished report, Psychol. Dept., Boston Univer., 1963.

Nakata, K. Synthesis and perception of nasal consonants. *J*. *acoust*. *Soc*. *Amer*., 1959, 31, 661-666.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word*, 1957, 13, 24-43.

Peterson, G. E. Vocal gestures. *Bell* *Lab*. *Rec*., 1951, 29, 500-503, 510. (a)

Peterson, G. E. The phonetic value of vowels. *Language*, 1951, 27, 541-553. (b)

Peterson, G. E. The information-bearing elements of speech. *J*. *acoust*. *Soc*. *Amer*., 1952, 24, 629-637.

Peterson, G. E. Phonetics, phonemics and pronunciation: spectrographic analysis. In E. Pulgram (Ed.), Applied linguistics in language teaching. *Georgetown* *Univer*. *Monogr*. *Lang*. *Ling*., 1954, No. 6. Pp. 7-18.

Peterson. G. E. Parameters of vowel quality. *J. Speech Hear. Res.*, 1961, 4, 10-29.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. *J. acoust. Soc. Amer.*, 1952, 24, 175-184.

Peterson, G. E., & Lehiste, Ilse. Duration of syllable nuclei in English. *J. acoust. Soc. Amer.*, 1960, 32, 693-703.

Pickett, J. M. Perception of compound consonants. *Lang. & Speech*, 1958, 1, 288-304.

Pickett, J. M., & Rubenstein, H. Perception of consonant voicing in noise. *Lang. & Speech*, 1960, 3, 155-163.

Pierce, J. R., & David, E. E. *Man's world of sound*. New York: Doubleday, 1958.

Potter, R. K., Kopp, G. A., & Green, H. C. *Visible speech*. New York: Van Nostrand, 1947.

Reeds, J. A., & Wang, W. S-Y. The perception of stops after /s/. *Phonetica*, 1961, 6, 78-81.

Rigault, A. Role de la fréquence de l'intensité et de la durée vocalique dans la perception de l'accent en français. In A. Sovijarvi & P. Aalto (Eds.), *Proced. 4th Int. Cong. Phon. Sci.* The Hague: Mouton, 1962. Pp. 735-748.

Sacia, C. F., & Beck, C. J. The power of fundamental speech sounds. *Bell Syst. tech. J.*, 1926, 5, 393-403.

Schatz, C. The role of context in the perception of stops. *Language*, 1954, 30, 47-56.

Sharf, D. J. Duration of post-stress intervocalic stops and preceding vowels. *Lang. & Speech*, 1962, 5, 26-30.

Shearme, J. N., & Holmes, J. N. An experimental study of the classification of sounds in continuous speech according to their distribution in the formant 1 --formant 2 plane. In A. Sovijarvi & P. Aalto (Eds.), *Proced. 4th Int. Cong. Phon. Sci.* The Hague: Mouton, 1962. Pp. 234-240.

Smith, S. Vocalization and added nasal resonance. *Folia Phoniat.*, 1951, 3, 165-169.

Söderberg, C-G. A typological study on the phonetic structure of English words with an instrumental-phonetic excursus on English stress. *Travaux L'Inst. Phonet. Lund*, 1959, No. 1

Stevens, K. N., & House, A. S. Studies of formant transitions using a vocal tract analog. *J. acoust. Soc. Amer.*, 1956, 28, 578-585.

Stevens, K. N., & House, A. S. Perturbation of vowel articulations by consonantal context: an acoustical study. *J. Speech Hear. Res.*, 1963, 6, 111-128.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.* 1958, 54, 177-196.

Stevens, S. S., & Volkman, J. The relation of pitch to frequency: a revised scale. *Amer. J. Psychol.*, 1940, 53, 329-353.

Strevens, P. Spectra of fricative noise in human speech. *Lang. & Speech*, 1960, 3, 32-49.

Wang, W. S-Y. Transition and release as perceptual cues for final plosives. *J. Speech Hear. Res.*, 1959, 2, 66-73.

Wang, W. S-Y. Stress in English. *Lang. Learn.*, 1962, 12, 69-77.

Wang, W. S-Y., & Filmore, C. J. Intrinsic cues and consonant perception. *J. Speech Hear. Res.*, 1961, 4, 130-136.

Footnotes

1. For a general introduction to phonetics, the widely used text by Heffner (1952) may be recommended. The traditional methods of linguistics for delimiting an inventory of basic speech sounds for a language are described rigorously by Harris (1951). New methods and criteria for their adequacy have come from recent theoretical developments in linguistics; see Jakobson, Fant and Halle (1952), Halle (1954), Chomsky (1957), Halle (1962). An introduction to the acoustics of speech is provided by Ladefoged's readable little book (1962); an advanced and comprehensive treatment will be found in the important treatise by Fant (1960). Traditional methods of experimental psychology for the measurement of discriminative behavior are reviewed by Stevens (1958), and their applications to speech discrimination in particular are reviewed by Licklider & Miller (1951). Recent elaborations and innovations in these methods are described in Luce, Bush & Gallanter (1963). Modern techniques for the analysis of nonverbal human and infra-human behavior are being applied to this area (Mostofsky, 1964).

2. Joos (1948) has presented the hypothesis that the listener normalizes the different acoustic vowel diagrams of different speakers by reference to an invariant neural map of the articulatory vowel diagram.

3. Comparable findings have been reported, for example, by Peterson (1961), who presents the measurements for individual speakers. The parameters of fundamental frequency and formant amplitude are discussed below.

4. In the latter study, 1263 monosyllables were uttered in the sentence frame "Say the word. . . again." Primary stress and a change from highest to lowest intonation occurred on the fourth word; measurements were taken at the peak of the intonation contour. These data are displaced in Fig. 22 above the average measurements for isolated monosyllables.

5. See the section on vowel nasalization below. Additional measurements of formant bandwidths are discussed in Stevens and House (1963), Dunn (1961) and Fant (1960, pp. 125 ff.).

6. A more recent study by Peterson (1961), which presents measurements of this parameter for individual speakers, found that the higher formants had more energy than reported in the previous study.

7. Smith (1951); Delattre (1954); House and Stevens (1956); Hattori, Yamamoto and Fujimura (1958); Fant (1960); Dickson (1962).

8. These views are discussed in Delattre (1958b), Liberman (1957), and Lane (1964).

9. The frequency distributions for [b, d, g] appear in Fig. 54.

10. It will be observed that [k, g] before these vowels have a locus close to that of the bilabials. When Halle, Hughes and Radley (1957) deleted the bursts of velar plosives in postvocalic position, [uk] (but not, for example, [Ik]) was poorly identified and often judged as [up]. These and other considerations reviewed by Delattre (1958a) indicate that the cues for velar articulation preceding or following [ɔ, o, u] are mainly or entirely in the burst following the closure rather than in the formant transitions. Also see Lehiste (1961a)

11. However, Delattre (1958b) has stated recently that the hypothesis is no longer current which assigns to the locus the resonant frequency of the oral cavity before the burst of the plosive.

12. Stevens and House (1956) employed an electrical analog of the vocal tract to synthesize the plosives and to examine their associated transitions. In general, the results support the locus hypothesis but require a somewhat weaker statement of the invariances associated with place of articulation. Durand (1954, 1956) provides further evidence for the locus hypothesis.

13. Pickett and Rubenstein (1960) have shown that the voicelessness of [t] is more readily detected in low-frequency masking noise than that of [p]. Since the alveolar locus is considerably higher in frequency than the bilabial locus, the initial segment of the F2 transition, which provides the aspiration cue to voicelessness, may have been more audible above low frequency masking noise in the case of [t] than in the case of

[p]. Further effects of masking on the identification of consonants are described by Pickett (1958).

14. The findings of Lotz and his collaborators (1960) in a comparable experiment underscore the necessity of keeping the discriminative history of the listener in mind when making predictions about his responses to speech. These investigators showed that Puerto Rican, Hungarian and Thai listeners, in whose languages there are unaspirated voiceless plosives, judged that the truncated syllables, such as "(s)pat", began with a voiceless plosive.

15. Yet another acoustic cue for the voiced-voiceless distinction should be mentioned. The preceding observations have centered mainly on plosives in prevocalic position. Some experiments with natural language, described later in this section, are concerned with plosives in postvocalic position. Here we may note that, in intervocalic position, voiced and voiceless plosives are distinguished by the duration of the silent interval ("closure"). Lisker (1957b) spliced segments of blank tape into the closure interval between the transitions of the preceding and following vowels in the words "rupee" and "ruby" and found that in both cases the plosive was identified as voiced [b] at short durations (less than 60 msec) and voiceless [p] at longer durations (greater than 130 msec).

16. Hogan (1963). The preceding findings, for the voiceless fricatives are substantially in agreement with the spectrographic measurements of Strevens (1960) for English and Jassem (1962) for Polish as well as the results of spectral analyses by Mártony et al. (1962) and Jassem (1963) for Swedish. The latter report, however, highlights the marked differences between speakers in the spectral properties of their fricatives.

17. Hughes & Halle (1956) also showed that spectral distribution is a sufficient cue for the identification of the fricatives [f], [s] and [ʃ]. They isolated the friction segments of recorded syllables beginning with these consonants and presented them to listeners, who were quite consistent in identifying them.

18. Strevens (1960); [s] is intermediate followed by [ʃ] and, highest, [h]. This order agrees with the measurements of Sacia and Beck (1926) but a modern systematic study of fricative intensities is not available.

19. Acoustic measurements of the nasal consonants are reported for Russian by Fant (1960), for Polish by Jassem (1962), for Japanese by Hattori, Yamamoto and Fujimura (1958), for French by Delattre (1954) and the parameter values of English nasals in various studies have been tabulated by Beresford (1963).

20. Cf. Fig. 51; Fant (1960); House (1957).

21. Wang and Filmore (1961) obtained similar results.

22. House (1957); Hecker (1962); Malecot (1956); Nakata (1959).

23. The discussion of the acoustic properties of [w, j, r, l] draws extensively from a systematic spectrographic investigation by Lehiste (1962b).

24. In his survey of the acoustical indices of speech, Delattre (1958b) cites only one experiment on the prosodic features. Magdics (1963) reviews studies of intonation here and abroad during the past decade. Hadding-Koch (1961) has conducted a major acoustic-phonetic study of Swedish intonation but her findings are largely pertinent to the problem of stress, discussed below, and to the perception of affect in speech.

25. Dreher (1951) and Goodman (1952) also present spectrographic analyses of English intonation.

26. Also see Goldman-Eisler (1961a). Maclay and Osgood (1959) showed that pauses occurred more frequently before lexical words than function words.

27. Several other phonetic and phonemic issues have been subsumed at one time or another under the heading of juncture. Lehiste (1960) reviews these issues, which have little to contribute to the present discussion of acoustic invariances.

28. In a comparable experiment which was performed subsequently (Fry, 1958), the ranges of duration and intensity effected a 70 and a 42 percent increase in stress judgments, respectively.

29. Söderberg (1959), Bolinger & Gerstman (1957) and Bolinger (1958b) present additional evidence and arguments for the conclusion that intensity makes a negligible contribution to the discrimination of English stress. Contrary to these several findings, some scientists believe that intensity is indirectly a cue for the perception of stress, and that its effects are mediated by covert articulation. Their arguments are evaluated in Lane (1964).

30. The importance of fundamental frequency as a cue for the discrimination of stress has been documented for Polish by Jassem (1959), for Swedish by Malmberg (1956) and Hadding-Koch (1961), for Serbo-Croatian by Lehiste (1961b) and for an Aztec language by Bascom (1959).

Table I

Acoustic parameters of American English vowels, uttered in the context h-d by men, women, and children. From Peterson and Barney, 1952.

		i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɝ
Fundamental frequencies (cps)	M	136	135	130	127	124	129	137	141	130	133
	W	235	232	223	210	212	216	232	231	221	218
	Ch	272	269	260	251	256	263	276	274	261	261
Formant frequencies (cps)											
F_1	M	270	390	530	660	730	570	440	300	640	490
	W	310	430	610	860	850	590	470	370	760	500
	Ch	370	530	690	1010	1030	680	560	430	850	560
F_2	M	2290	1990	1840	1720	1090	840	1020	870	1190	1350
	W	2790	2480	2330	2050	1220	920	1160	950	1400	1640
	Ch	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
F_3	M	3010	2550	2480	2410	2440	2410	2240	2240	2390	1690
	W	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
	Ch	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160
Formant amplitudes (db)	L_1	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
	L_2	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
	L_3	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

Table II

Average formant frequencies of American English vowels, as measured at the formant targets in several consonantal environments. From Lehiste and Peterson, 1961b.

Syllable nucleus	Averages for 1263 CNC words spoken by GEP			Averages for 350 CNC words spoken by 5 speakers		
	F_1	F_2	F_3	F_1	F_2	F_3
i	315	2200	2700	320	2205	2800
I	415	1750	2470	410	1755	2415
e^{I}	360	2015	2510	335	2105	2630
ɛ	570	1610	2465	540	1705	2415
æ	640	1570	2460	625	1740	2415
ə	610	1185	2565	585	1155	2255
ɑ	645	1110	2540	665	1145	2520
ɔ	505	880	2525	590	985	2365
o^{υ}	495	960	2495	435	905	2435
U	450	980	2360	400	1015	2090
u	355	895	2240	350	845	2105
ɑʊ	655-510	1255- 910	2520-2415	655-415	1235- 870	2215-2225
ɑI	665-485	1200-1790	2540-2450	700-375	1315-1975	2360-2585
ɔI	510-505	900-1610	2510-2425	550-370	950-1830	2255-2485
ɝ	475	1245	1680	430	1255	1575

Table III

Formant frequencies of synthetic vowel stimuli in the experiment by Fry *et al*., 1962. (Cf. Fig. 21)

STIMULUS NUMBER	FIRST FORMANT	SECOND FORMANT
1	330	1980
2	380	1970
3	410	1960
4	460	1930
5	490	1910
6	500	1890
7	550	1880
8	580	1860
9	650	1860
10	700	1820
11	780	1820
12	830	1780
13	890	1760

Table IV

The influence of initial and final consonants on the average amplitudes of American English vowels. From Lehiste and Peterson, 1959b.

Amplitude of vowels	Initial consonant	Amplitude of vowels	Final consonant
80.5	r	80.3	b
80.2	l	80.1	g
79.9	v	80.0	r
79.7	b d g	79.9	l
79.5	w	79.8	d v
79.4	m	79.7	dʒ
79.2	dʒ	79.4	ʃ
79.1	ð n	79.2	z
79.0	y f	79.1	ŋ
78.6	w h	79.0	n
78.4	z	78.9	m f
78.3	θ s	78.8	s t
78.2	ʃ	78.7	ð ʒ
78.1	h	78.5	θ
77.9	k	78.0	k
77.7	t tʃ	77.5	tʃ
77.3	p	77.4	p

Table V

The effects of the frequency of the burst and the formant transitions of the vowel on the identification of plosives. The findings obtained with tape-splicing of recorded utterances are compared to those obtained with the synthesis of plosives from hand-painted spectrograms. From Schatz, 1954.

In actual speech				In synthetic speech			
	syllables formed				syllables formed		
burst	with i	a	u	burst	with i	a	u
ki	ki	ta	pu	2880	ki	ta	pu, tu
ka	pi	ka	pu	1440	pi	ka	pu
ku	pi, ki	pa	ku	720	pi, ki	pa	ku

Table VI

The effects of interchanging the nasal murmurs and vowel-formant transitions of [mæ, næ, ŋæ] on the identification of nasal consonants. From Malecot, 1956.

	Consonant in Initial Position	% Judgments				Consonant in Terminal Position	% Judgments		
No.		m	n	ŋ	No.		m	n	ŋ
70	m(æ) + (m)æ	100	0	0	79	æ(m) + (æ)m	100	0	0
71	m(æ) + (n)æ	16	64	20	80	æ(n) + (æ)m	96	4	0
72	m(æ) + (ŋ)æ	16	16	68	81	æ(ŋ) + (æ)m	96	0	4
73	n(æ) + (m)æ	96	4	0	82	æ(m) + (æ)n	64	36	0
74	n(æ) + (n)æ	0	96	4	83	æ(n) + (æ)n	0	100	0
75	n(æ) + (ŋ)æ	0	12	88	84	æ(ŋ) + (æ)n	0	52	48
76	ŋ(æ) + (m)æ	100	0	0	85	æ(m) + (æ)ŋ	56	40	4
77	ŋ(æ) + (n)æ	8	72	20	86	æ(n) + (æ)ŋ	4	96	0
78	ŋ(æ) + (ŋ)æ	0	12	88	87	æ(ŋ) + (æ)ŋ	0	20	80

Table VII

Average formant frequencies for repeated utterances by five American speakers of [w, j, l, r] in initial position before various syllable nuclei. After Lehiste, 1962b.

	F1	F2	F3
[w]	305	630	2180
[j]	265	1985	2850
[l]	295	950	2610
[r]	280	920	1350

Table VIII

For each of 16 American speakers who uttered 38-50 pairs of bisyllabic words with contrastive stress (e.g., PERmit, perMIT) the table shows the number of words in which the frequency, amplitude and duration of the stressed syllable was higher than that of the unstressed syllable in the same word (II, III, IV) or in the other word of the stress pair (V, VI, VII). From Lieberman, 1960.

		Word basis			Pair basis		
Speaker No.	No. of words I	Freq. II	Amp. III	Duration IV	Freq. V	Amp. VI	Duration VII
1	50	48	43	35	43	45	40
2	50	50	45	32	48	45	38
3	43	40	39	31	39	39	35
4	44	42	34	30	29	30	29
5	46	40	38	30	37	42	30
6	43	38	39	27	23	40	27
7	50	40	41	39	34	45	40
8	48	39	40	30	31	42	39
9	41	36	38	25	23	39	29
10	46	43	43	37	36	44	38
11	44	35	37	32	32	36	40
12	42	38	33	30	30	40	30
13	41	36	41	25	25	40	23
14	38	34	34	26	25	37	22
15	39	36	35	25	24	35	24
16	41	36	34	32	24	32	33

Table IX

The effects of varying the acoustic parameters of the first syllable of "papa" on the perception of which syllable is stressed. After Rigault (1962).

Word No.	Parameters of the initial syllable			Stress locus (responses)		Stress locus (percent)	
	Frequency	Intensity	Duration	First syllable	Second syllable	First syllable	Second syllable
5	high	strong	long	40	0	100%	0%
6	high	strong	short	39	1	97.5%	2.5%
7	high	weak	long	35	5	87.5%	12.5%
8	high	weak	short	28	12	70%	30%
1	low	strong	long	16	24	40%	60%
4	low	weak	short	9	31	22.5%	78.5%
2	low	strong	short	6	34	15%	85%
3	low	weak	long	6	34	15%	85%

Table X

The difference in the percent of times that the second syllable was judged stressed in pairwise comparisons of the eight stimuli described in Table IX. (frequency: F=high, f=low; intensity: I=strong, i=weak; duration: D=long, d=short.) After Rigault, 1962.

Order	Difference in percent of stress judgments	Words compared	Parameters of the initial syllable of each word	Difference in parameters within pairs		
1.	-85%	5--2	F I D -- f I d	f-		d-
2.	-85%	5--3	F I D -- f i D	f-	i-	
3.	-82.5%	6--2	F I d -- f I d	f-		
4.	-82.5%	6--3	F I d -- f i D	f-	i-	d+
5.	-77.5%	5--4	F I D -- f i d	f-	i-	d-
6.	-75%	6--4	F I d -- f i d	f-	i-	
7.	-72.5%	7--2	F i D -- f I d	f-	i+	d-
8.	-70%	7--3	F i D -- f i D	f-		
9.	-65%	7--4	F i D -- f i d	f-		d-
10.	-60%	5--1	F I D -- f I D	f-		
11.	-57.5%	6--1	F I d -- f I D	f-		d+
12.	-55%	8--2	F i d -- f I d	f-	i+	
13.	-55%	8--3	F i d -- f i D	f-		d+
14.	-47.5%	8--4	F i d -- f i d	f-		
15.	-47.5%	7--1	F i D -- f I D	f-	i+	
16.	-30%	8--1	F i d -- f I D	f-	i+	d+
17.	-30%	5--8	F I D -- F i d		i-	d-
18.	-27.5%	6--8	f I D -- F i d		i-	
19.	-25%	1--3	F I d -- f i D		i-	
20.	-25%	1--2	f I D -- f I d			d-
21.	-17.5%	7--8	F i D -- F i d			d-
22.	-17.5%	1--4	f I D -- f i d		i-	d-
23.	-12.5%	5--7	F I D -- F i D		i-	
24.	-10%	6--7	F I d -- F i D		i-	d+
25.	-2.5%	5--6	F I D -- F I d			d-
26.	0	2--3	f I d -- f i D		i-	d+
27.	+ 7.5%	3--4	f i D -- f i d			d-
28.	+ 7.5%	2--4	f I d -- f i d		i-	

Figure Captions

Fig. 1. A spectrogram of the utterance "Bell Telephone Laboratories." The major resonances of the speech sounds are displayed as a function of time. From Peterson (1951a).

Fig. 2. The upper portions of the patterns show spectrograms of several sustained vowels, which are identified by English and phonetic spellings. The lower portions display the amplitudes of each of the component frequencies of the speech sounds at one point in time. The frequency scale on the vertical axis has been inverted. From Peterson (1951a).

Fig. 3. The vowel utterances (phones) whose spectrograms and amplitude sections are shown in Fig. 2 are represented here by points whose coordinates are the frequencies of the first and second formants of each vowel. The origin of the acoustic vowel diagram has been placed at the upper right so that the vowel points have approximately the same spatial arrangement as in an articulatory vowel diagram. After Peterson (1951a).

Fig. 4. The formant frequencies of English vowels spoken by a man, a woman and a child. From Peterson (1951a).

Fig. 5. An acoustic vowel diagram for English vowels spoken by a man, a woman, and a child. The frequencies of the first and second formant of each vowel phone were expressed in units of pitch (mels) and then normalized with respect to the pitch of the third formant of that phone. From Peterson (1951b).

Fig. 6. The formant frequencies of English vowels uttered on several different days by each of two speakers (A, B). From Peterson (1954).

Fig. 7. Frequency distributions of the second formant of [æ] and of [ɛ] in repeated utterances by one speaker. From Lisker (1948).

Fig. 8. The formant frequencies of English vowels, uttered repeatedly by men and children and identified unanimously by a group of listeners. From Peterson and Barney (1952).

Fig. 9. Spectrograms of three diphthongs spoken by an American. From Pierce and David (1958).

Fig. 10. The median formant frequencies of six diphthongs, uttered by 20 General American speakers. The direction of change of the formants during five consecutive time samples is shown by the arrows. The dashed lines encompass most of the distribution of points for each diphthong represented by the five medians (dots). From Holbrook & Fairbanks (1962).

Fig. 11. Transitions of vowel formants in consonantal environments. Left: Formant frequencies, in consecutive time samples, for the vowel phone [ʊ], determined by "analysis by synthesis". The vertical lines represent the boundaries of that vowel in the utterance [hə'dʊd], which were determined by visual examination of spectrograms. Right: Approximations to the spectral data obtained by fitting parabolic curves to the plotted points within the vowel boundaries. After House *et al*. (1963).

Fig. 12. Effects of consonantal environments on the formant frequencies of vowels. Each point represents the average formant frequencies of a vowel produced repeatedly in one of 14 consonantal environments by one of three speakers. The contours are redrawn from Fig. 8, where they enclosed most of the points representing utterances of the vowels in a constant (neutral) environment. From Stevens and House (1963).

Fig. 13. Effects of connected speech on the formant frequencies of the vowel [i]. The changing formants of the vowel during each of its utterances are indicated by the directional tracks (cf. Fig. 10). The dashed line encompasses the minimum area which contains part of each track. The cross indicates the formant frequencies of [i] when it was uttered in an isolated monosyllable by the same speaker. From Shearme and Holmes (1962).

Fig. 14. Effects of connected speech on the formant frequencies of eight vowels in English. Directional tracks were drawn representing the occurrences of each

vowel during discourse by one British speaker (cf. Fig. 13). The minimum area containing part of each track for the indicated vowel is shown. Also shown are the formant frequencies of the vowels when uttered in isolated monosyllables by the same speaker. From Shearme and Holmes (1962).

Fig. 15. Successive reductions in the areas in the formant one--formant two plane occupied by General American vowels, based on the discriminative behavior of listeners. From Fairbanks and Grubb (1961).

Fig. 16. Identification of vowels when all frequencies above (low pass) or below (high pass) the indicated frequency were eliminated (or increasingly attenuated). The curves showing the highest intelligibility were obtained by Steinberg with General American long vowels in consonantal environments and ideal filters (reported in Fletcher, 1953). The intermediate set of articulation functions was obtained with 1500 msec vowel segments extracted from their environments and more gradual filtering (40 db/ octave; Lehiste and Peterson, 1959a). The lowest articulation scores (filled symbols) were reported in an article by Carterette and Møller (1962) in which this figure appeared. Segments of Swedish vowels, 500 msec long, were extracted from recorded monosyllables and passed through ideal filters.

Fig. 17. Identification of vowels, uttered by General American speakers, when a 670-cycle low pass filter was used to eliminate the second formant. The heavy lines indicate frequent confusions, the light lines less frequent confusions, and the broken lines infrequent confusions. From Miller (1956).

Fig. 18. "Diagram of pattern playback. The tone wheel has 50 circular sound tracks which yield pure tones at 120 cps intervals from 120 to 6000 cps when the tone wheel is rotated at 1800 rpm by a synchronous motor. Light (from the lamp at the extreme left) is modulated by the tone wheel and directed onto the spectrogram in such a way that the spectrographic patterns will transmit (or reflect) to a photocell just those portions of the light which carry the frequencies corresponding to the pattern." From Cooper (1953, p. 46).

Fig. 19. Identification of seven disparate synthetic vowels represented by points in an acoustic vowel diagram. Within each circle, the value given above the line is the percentage of correct identifications and the value below the line is the percentage of correct identifications plus the percentage of identifications which were incorrect to the extent of only one step in the vowel chart. From Delattre et al. (1952).

Fig. 20. Identification of 16 synthetic vowels prepared to approximate the cardinal vowels of the I. P. A. Correct identifications are indicated as in Fig. 19. From Delattre et al. (1952).

Fig. 21. Identifications by American listeners of 13 synthetic vowels whose parameters are listed in Table III. From Fry et al. (1962).

Fig. 22. Fundamental frequency of vowels and diphthongs in General American English. The dashed lines represent average values reported by Peterson and Barney (1952; cf. Table I). The solid lines represent measurements, taken at the peak of an intonation contour, reported in an article by Lehiste and Peterson (1961a) in which this figure appeared.

Fig. 23. Fundamental frequency as a cue for discriminating between words. Spectrograms of the lower harmonics of two series of Thai words. Within each series, the words are differentiated according to characteristic temporal patterns of fundamental frequency ("tones"). From Abramson (1962).

Fig. 24. Characteristic temporal patterns of fundamental frequency observed in single vowels in Thai. From Abramson (1962).

Fig. 25. Preferences of American listeners for synthetic vowels as a function of the bandwidths of the vowel formants. The bandwidths ranged from narrow (Set A: $B_1 = 40$, $B_2 = 50$, $B_3 = 60$ cps) to wide (Set G: $B_1 = 220$, $B_2 = 260$, $B_3 = 300$ cps). The response curves are plotted with overlapping abscissae. From House (1960).

Fig. 26. Average sound pressure levels of vowels and diphthongs spoken in isolation (top curve) and in monosyllables by General American talkers. From

Lehiste and Peterson (1959b).

Fig. 27. The durations of syllable nuclei (vowels and diphthongs) in General American English, averaged over their occurrences in consonantal environments. The solid curve is for a large set of monosyllables uttered by one speaker; the dotted curve represents the values for 30 minimal pairs uttered by five speakers. From Peterson and Lehiste (1960).

Fig. 28. The average durations of syllable nuclei in English as a function of the final consonant. The top curve represents the average durations for all long syllable nuclei (cf. Fig. 27) and the bottom curve the averages for four short syllable nuclei. From Peterson and Lehiste (1960).

Fig. 29. The magnitude of various influences on vowel duration in English. The durations of vowels in msec, averaged over repeated utterances by three American speakers, are shown partitioned according to two trinary and two binary variables. From House (1961).

Fig. 30. Vowel duration as a cue for discriminating between words. Spectrograms of the minimal pair of words in Thai [pæ̀] and [pæ̌æ̀] in a carrier sentence. The words are differentiated according to the duration of their vowel sounds. From Abramson (1962).

Fig. 31. The spectrum envelope of the combined output of the vocal tract and nasal tract for various degrees of coupling, as obtained from measurements on an electrical analog. The configuration of the vocal tract is appropriate for the production of the vowel /u/. From House and Stevens (1956).

Fig. 32. Judgments of nasality for five vowels synthesized on an electrical analog of the vocal tract with several degrees of coupling to an analog of the nasal tract. From House and Stevens (1956).

Fig. 33. Spectrograms of the voiceless and voiced plosives uttered by a Russian speaker. The accompanying spectral envelopes of the plosives were determined at the points in time indicated by the Roman numerals beneath the spectrograms. After Fant (1960).

Fig. 34. A spectrogram of the utterance [kɑp] by an American speaker, showing the plosive burst (*a*), the aspiration and formant transitions, and the beginning of voicing in the vowel (*b*). From Schatz (1954).

Fig. 35. Identification of English consonants in noise. The cues associated with the voicing of different consonants and those associated with their place of articulation emerge as two distinct homogeneous clusters. Voicing information is transmitted at signal-to-noise levels 18 db below those needed for place information. From Miller and Nicely (1955).

Fig. 36. Stimulus patterns used in determining the effect of the frequency of a burst of noise on the identification of the voiceless plosives. A: frequency positions of the 12 bursts of noise; B: frequency positions of the formants of the vowels with which the bursts were paired; C: one of the 84 syllables formed by pairing a burst of noise and a two-formant vowel. From Liberman, Delattre and Cooper (1952).

Fig. 37. Preferred identifications by 30 listeners of the stimuli of Fig. 36. The zones show the combinations of burst and vowel for which one of the three responses was dominant and indicate roughly the extent of dominance. After Liberman, Delattre and Cooper (1952).

Fig. 38. Stimulus patterns used in determining the effect of formant transitions on the identification of the plosives. A: the eleven values of second-formant transition used with one of the seven vowels; (the numbers indicate how many harmonics of the fundamental frequency, 120 cps, are crossed by the transition before arriving at the "steady state"). B: the two values of first-formant transition used with one of the seven vowels. C: formant frequencies of the seven synthetic vowels. From Liberman *et al*. (1954).

Fig. 39. Preferred identifications by 33 listeners of the stimuli illustrated in Fig. 38. The zones show the formant transitions for which [b], [d] or [g] was

dominant before each of the vowels, and they indicate by four levels of shading the approximate extent of the dominance. From Liberman et al. (1954).

Fig. 40. Formant transitions before seven vowels necessary to synthesize the labial, dental and velar plosives, respectively. The formant transitions associated with each place of articulation converge on an imaginary point, called the locus, at a particular frequency: 700, 1800, or 3000 cps, respectively. From Delattre (1962b).

Fig. 41. Hand-painted spectrographic patterns used to synthesize the most extremely voiced plosives (and vowels). From Liberman, Delattre and Cooper (1958).

Fig. 42. A series of spectrographic patterns for studying the discrimination of voiced and voiceless plosives as a function of the transition of the first formant. For each of the nine syllables shown in Fig. 41 (the illustration above is for /bæ/), the voice bar was removed and the first formant was increasingly cut back, in steps of 10 msec. From Liberman, Delattre and Cooper (1958).

Fig. 43. A series of spectrographic patterns for studying the discrimination of voiced and voiceless plosives as a function of the relative time of onset of the first formant. To form the seven stimuli of the experiment, the first formant was increasingly cut back, in steps of 10 msec, without altering its frequency at onset. From Liberman, Delattre and Cooper (1958).

Fig. 44. Spectrograms of six fricatives produced by a General American speaker. Top: [fa, θa, sa]; bottom: [va, ða, za].

Fig. 45. Spectrograms of four fricatives produced by a General American speaker. Top: [ʃa, ha]; bottom: [ʒa, ɦa].

Fig. 46. Spectrograms of two affricates produced by a General American speaker. Top: [tʃa]; bottom: [dʒa].

Fig. 47. A comparison of the spectral envelopes, measured by Hughes and Halle (1956), for the fricatives /f, s, ʃ/ spoken in different contexts by one American (dashed curves) and the corresponding spectral envelopes obtained as the output of

appropriate electrical networks (solid curves). From Heinz and Stevens (1961).

Fig. 48. Spectrographic patterns employed to synthesize four voiced fricatives. From Delattre (1962b).

Fig. 49. "Schematized description of a typical stimulus used in the identification test for fricative-vowel syllables. The solid bars in the upper part of the figure show how the formant frequencies change as a function of time in the vowel portion of the syllable, including typical transitions at the onset of buzz excitation and the steady formants corresponding to the vowel /ɑ/. The shaded area preceding the vowel depicts the fricative portion of a typical syllable. The lower part of the figure shows the form of the signals used to control the excitation and formant parameters in the synthesizer." From Heinz and Stevens (1961, p. 594).

Fig. 50. The identification of fricatives as a function of (a) the center frequency of the fricative noise, (b) the onset frequency of the following formants, and (c) the intensity ratio between the two. The contours show the range of the acoustic parameters within which the indicated fricative was identified greater than 75 percent (solid lines) or greater than 90 percent (dashed lines). From Heinz and Stevens (1961).

Fig. 51. Spectrograms of two nasal consonants uttered by a Russian speaker. The accompanying spectral envelopes of the nasal "murmurs" were determined at the points in time indicated by the Roman numerals beneath the spectrograms. After Fant (1960).

Fig. 52. X-ray tracings of the articulation of the nasal consonants [m] and [n]. From Fant (1960).

Fig. 53. Spectrographic patterns which, when converted into sound, yield the indicated stop and nasal consonants. The arrangement of the patterns according to the phonetic classifications of place and manner of articulation reveals the corresponding acoustic variables. From Liberman _et al_. (1956).

Fig. 54. Distributions of [m, n, ŋ] identifications as a function of the

formant transitions and steady state frequencies of the following vowel. The stimulus array is shown in Fig. 38. Distributions for the voiced plosives are also shown; these were summarized previously in Fig. 39. From Liberman _et_ _al_. (1954).

Fig. 55. Spectrograms of the semivowels and liquids produced by a General American speaker: [wɛ], [lɛ] / [rɛ], [jɛ].

Fig. 56. Spectrographic patterns that were used to study discriminations among consonants, semivowels and vowels. The first four patterns in each row illustrate how the tempo of the transitions was varied. A complete pattern (transition plus steady state vowel) is shown at the right for the longest duration of transition that was employed. From Liberman _et_ _al_. (1956).

Fig. 57. Identifications of the spectrographic patterns, illustrated in Fig. 56, as consonants, semivowels, and vowels, as a function of the durations of their formant transitions. From Liberman _et_ _al_. (1956).

Fig. 58. Spectrographic patterns which, when converted into sound, yield the indicated liquids and semivowels (cf. Fig. 55). From O'Connor _et_ _al_. (1957).

Fig. 59. A schematic representation of the spectrographic patterns used to study the identification of the semivowels and liquids as a function of the formant transitions and steady state frequencies in the following vowel. From O'Connor _et_ _al_. (1957).

Fig. 60. Identifications of the synthetic semivowels and liquids schematized in Fig. 59. Each frame shows the relative frequency of the indicated response as a function of the onset frequency of the second formant for each of the three third-formant conditions. From O'Connor _et_ _al_. (1957).

Fig. 61. Identification of the semivowels and liquids in intervocalic position as a function of the frequencies of their second and third formants and their vocalic environments. The contours show the range of formant frequencies within which the indicated identification was made by 50 percent or more of a

group of listeners. The crosses indicate stimuli for which the concensus of listeners was greatest. The dotted lines indicate the formant frequencies of the preceding and following synthetic vowel. From Lisker (1957a).

Fig. 62. Representative contours for declarative intonation in four languages, based on spectrographic analyses of conversation. The thickness of the lines indicates approximately the correlated changes in speech intensity. From Delattre (1962d).

Fig. 63. The correlation between vowel duration and vowel reduction. With decreasing duration, produced by decreasing stress, the frequency of the second formant increasingly "undershoots" the target value (dashed line), approximating instead the central frequency position of /œ/. After Lindblom (1963).

Fig. 64. The correlation between vowel duration and vowel reduction. With decreasing duration produced either by increased tempo (dots) or by decreased stress (curves), the second formant falls short of its target by roughly the same amount. This finding supports the assumption that vowel reduction is associated primarily with vowel length and only indirectly with stress. From Lindblom (1963).

Fig. 65. The relative frequency with which each of five bisyllabic words was judged to be stressed on the first syllable ("noun judgments") as a function of the ratio of the duration (left) or intensity (right) of the vowel in the first syllable to that in the second. From Fry (1955).

Fig. 66. The effect of varying the location of an intonation peak (top), in an otherwise constant synthetic sentence, on the syllable which is judged stressed (bottom). From Garding and Gerstman (1960).

Fig. 67. The relative frequency with which synthetic syllabic words were judged to be stressed on the first syllable ("noun judgments"), as a function of the intonation contour in the words. From Fry (1958).

PRECEDING PAGES MISSING Fig I + II

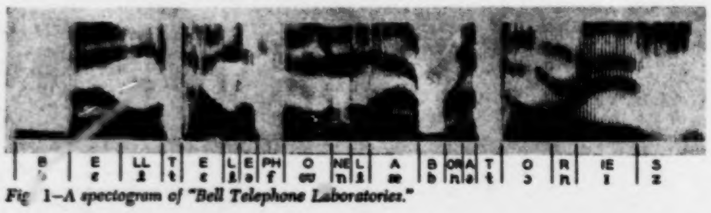


Fig. 1–A spectogram of "Bell Telephone Laboratories."

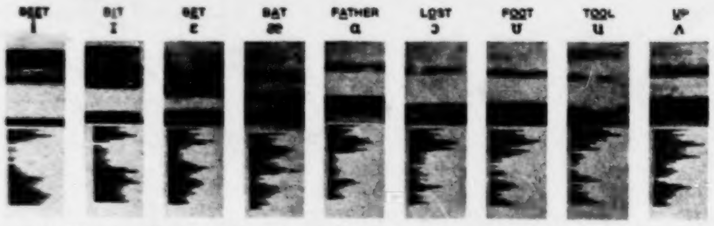


Fig. 2–The upper portions of the patterns show spectrograms of a number of sustained vowels. A section of each vowel is shown immediately below each spectrogram.

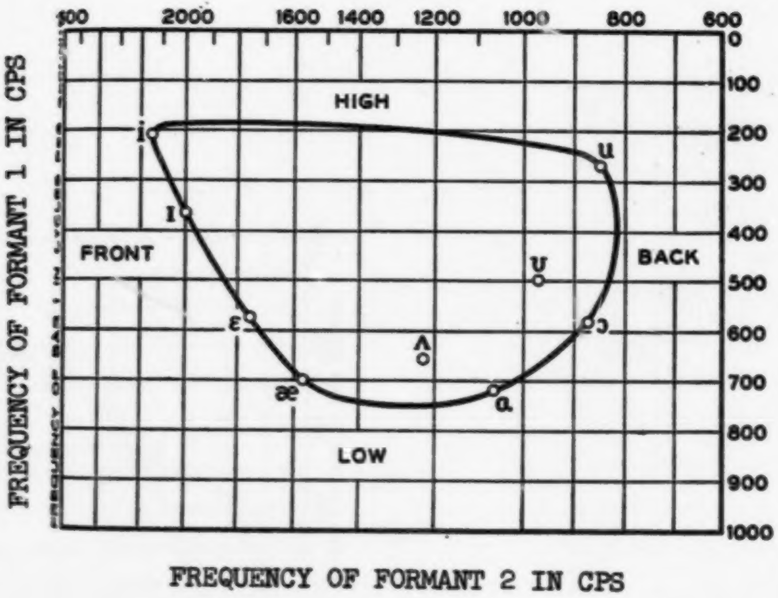


Fig. 3

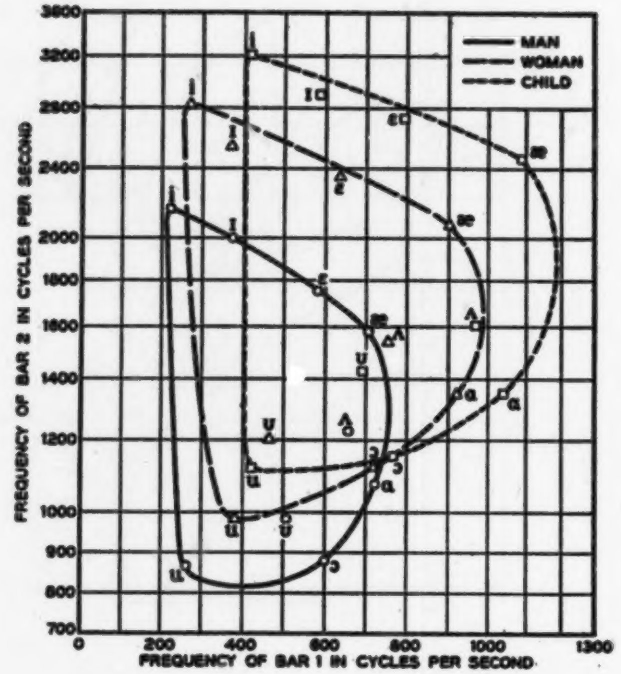


Fig. 4

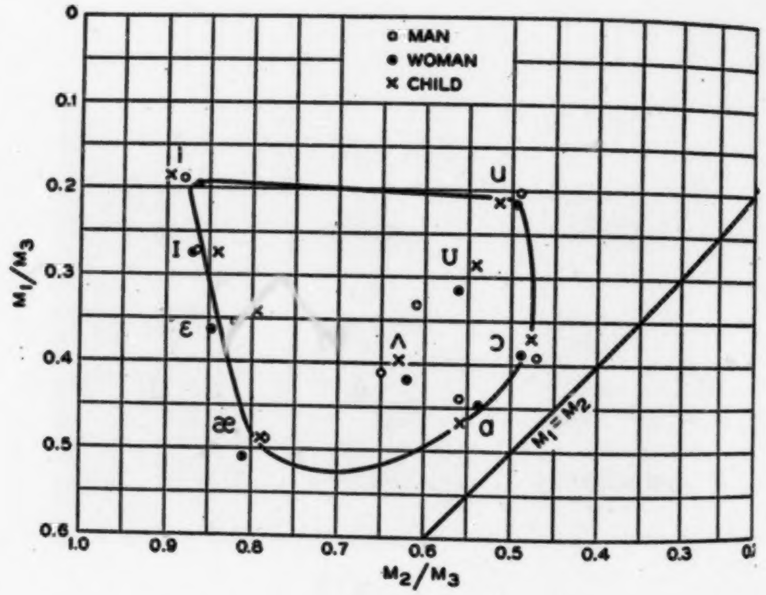


Fig. 5

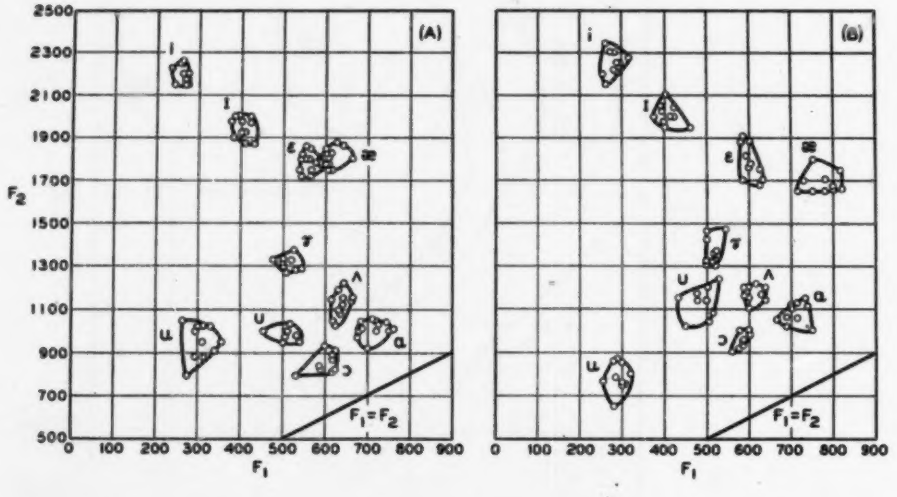


Fig. 6

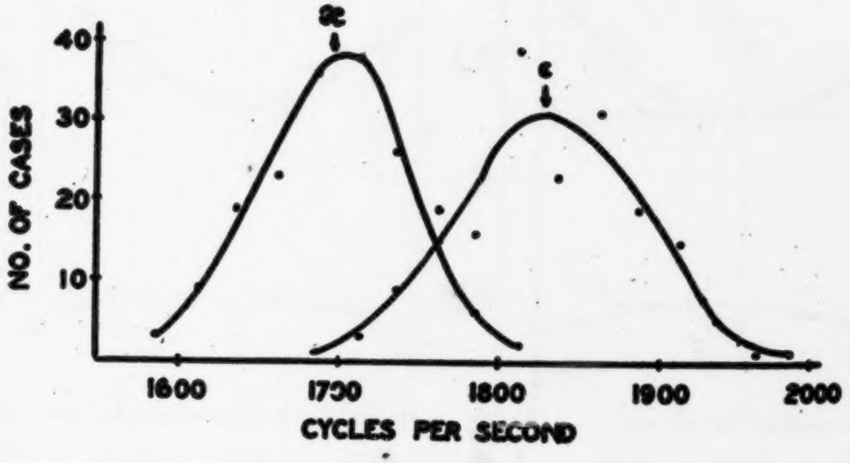


Fig. 7

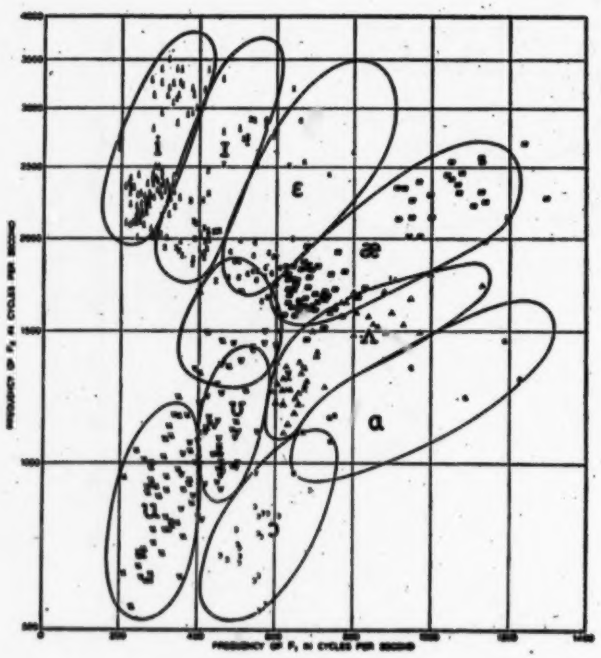


Fig. 8

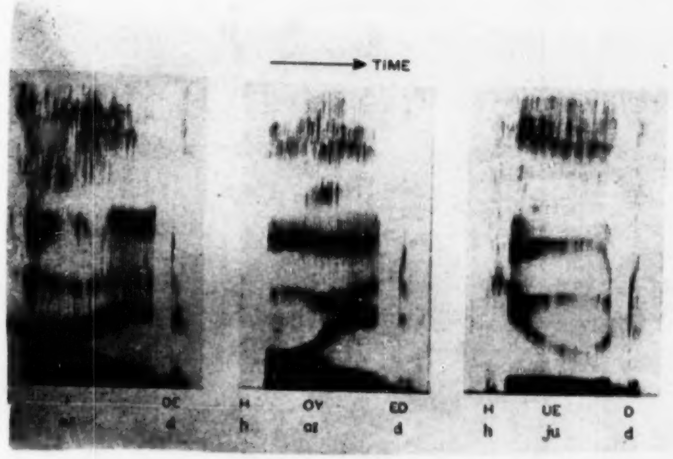


Fig. 9

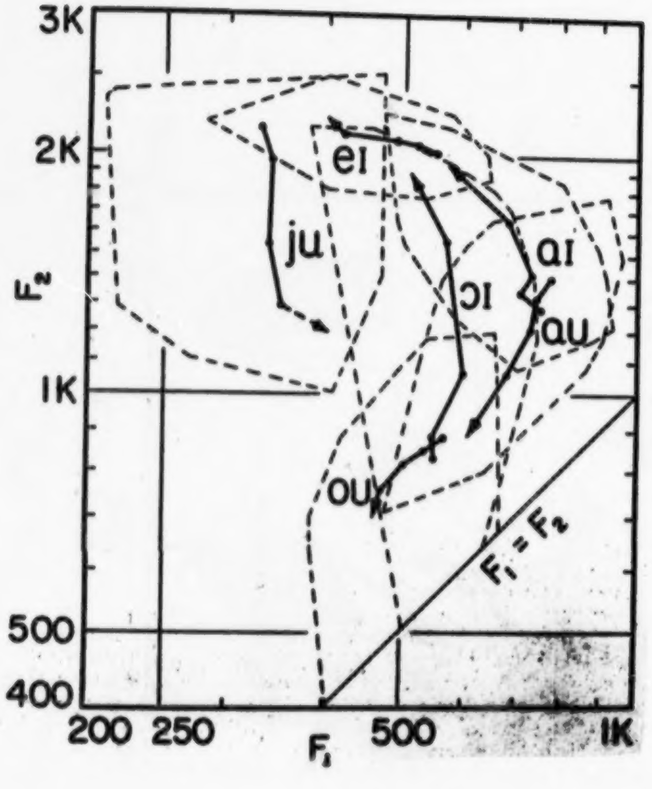


Fig. 10

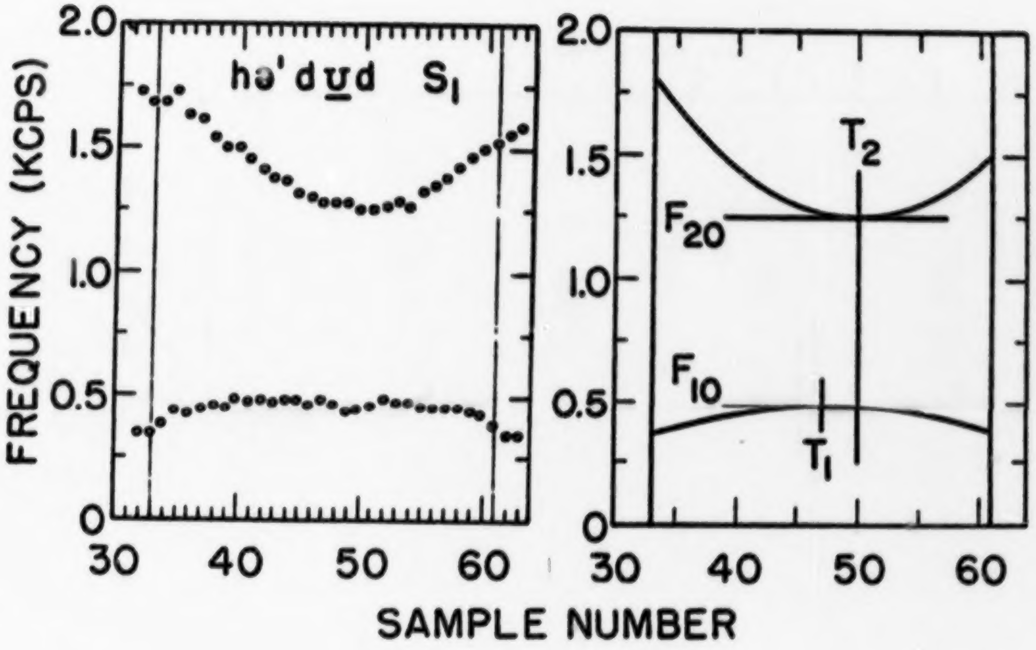


Fig. 11

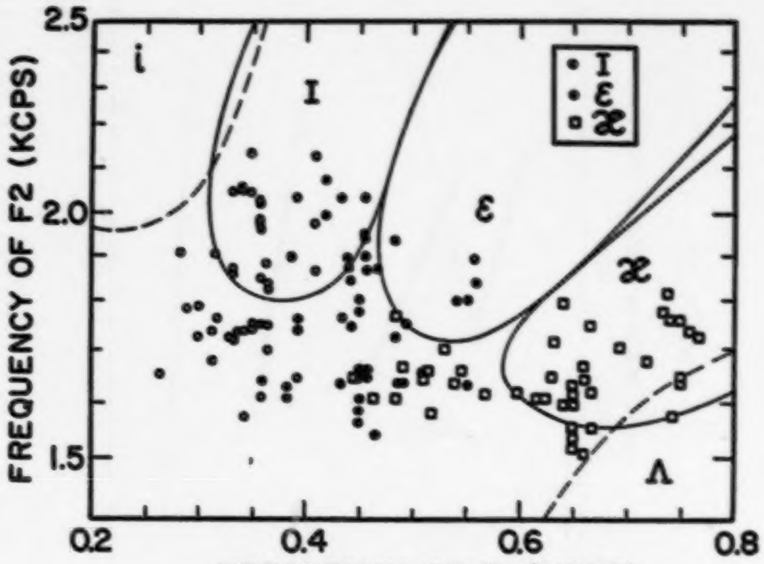


Fig. 12

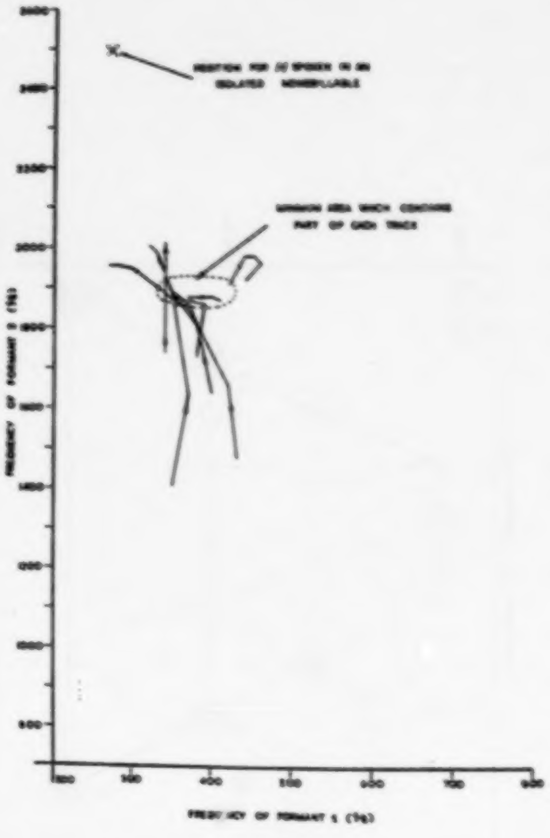


Fig. 13

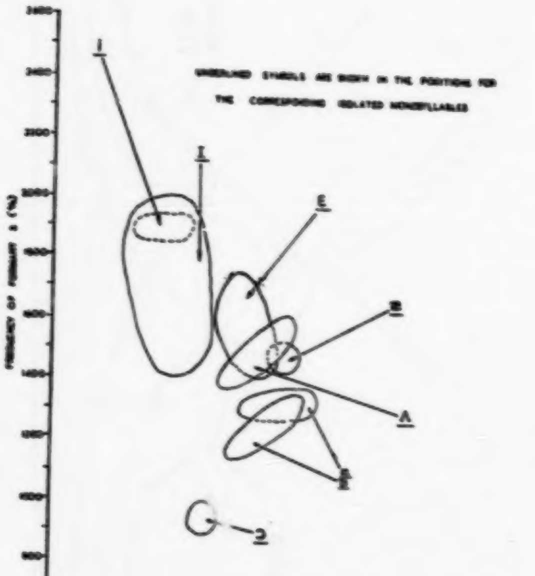


Fig. 14

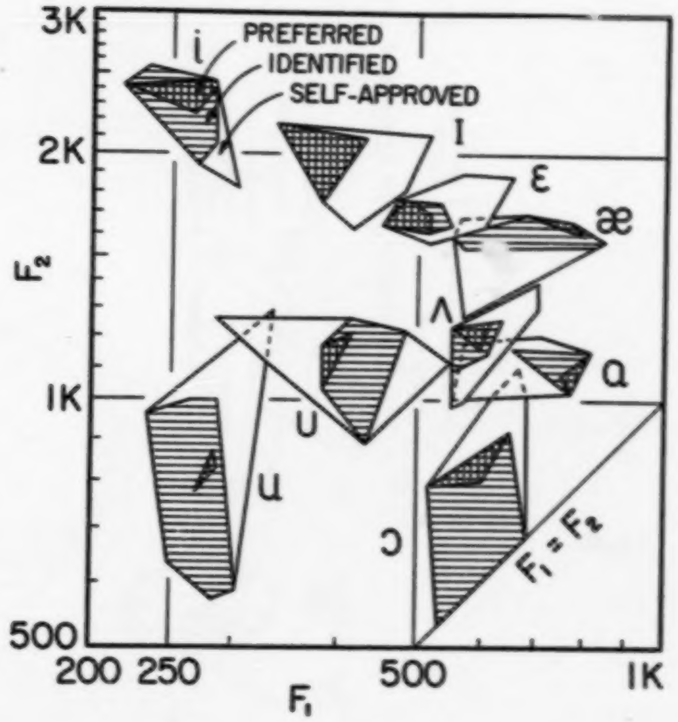


Fig. 15

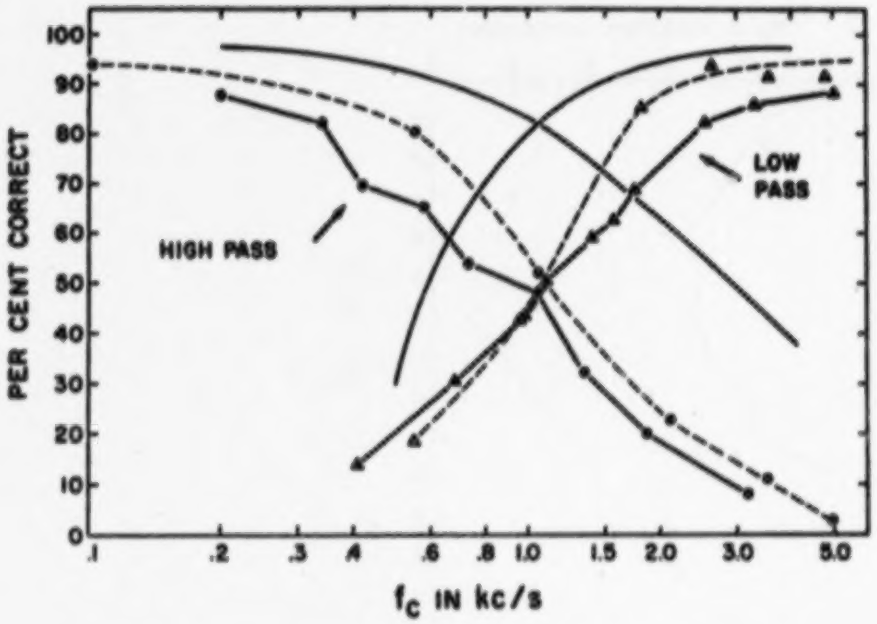


Fig. 16

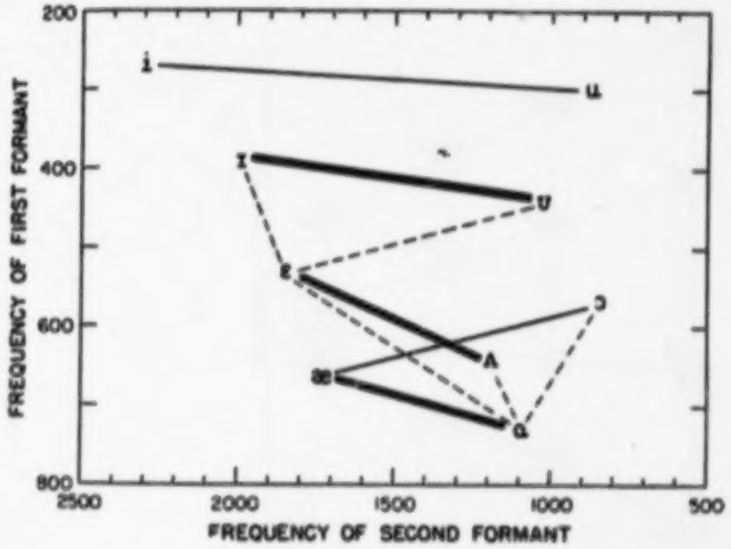


Fig. 17

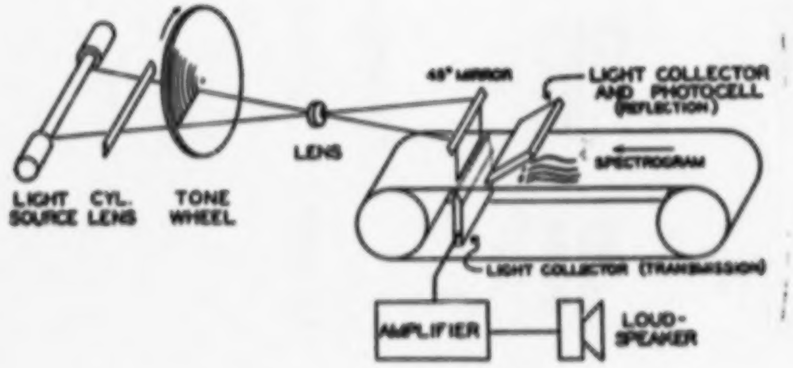

LIGHT SOURCE
CYL. LENS
TONE WHEEL
LENS
45° MIRROR
LIGHT COLLECTOR AND PHOTOCELL (REFLECTION)
SPECTROGRAM
LIGHT COLLECTOR (TRANSMISSION)
AMPLIFIER
LOUD-SPEAKER

Fig. 18

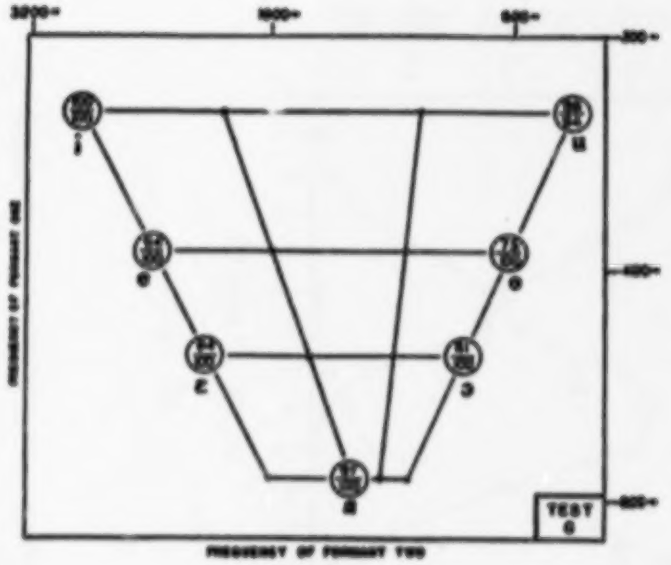


Fig. 19

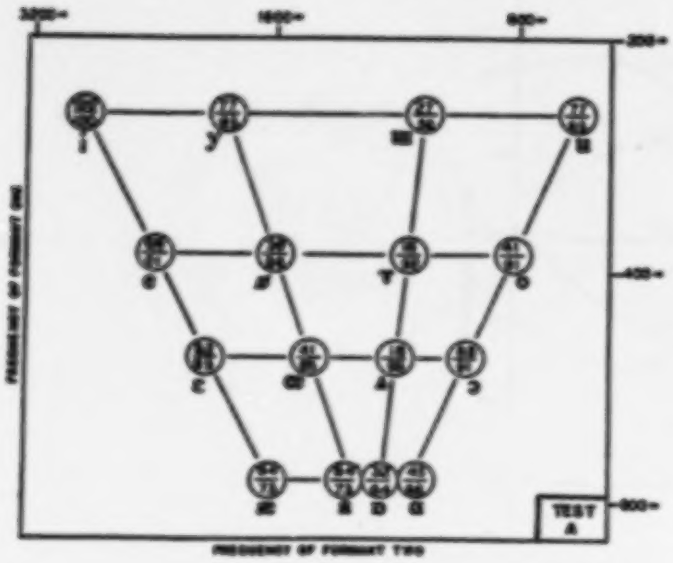


Fig. 20

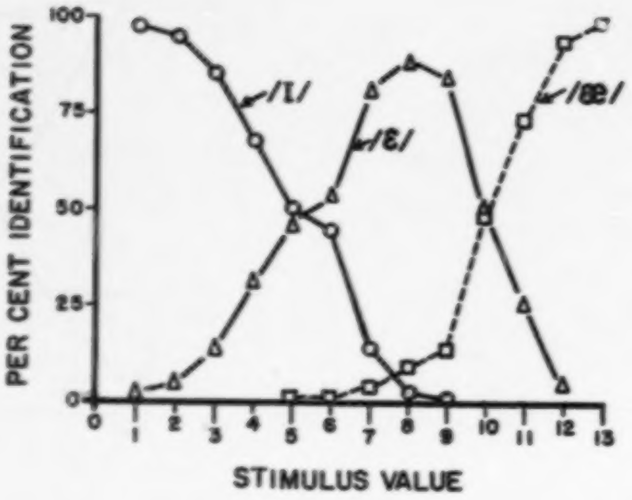


Fig. 21

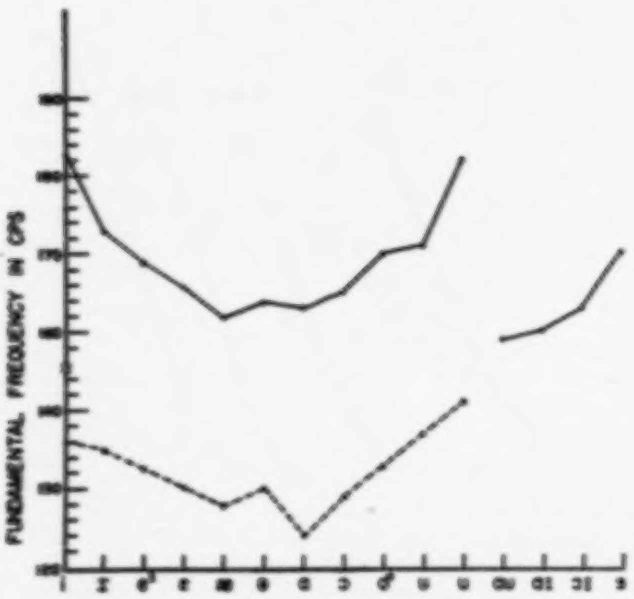


Fig. 22

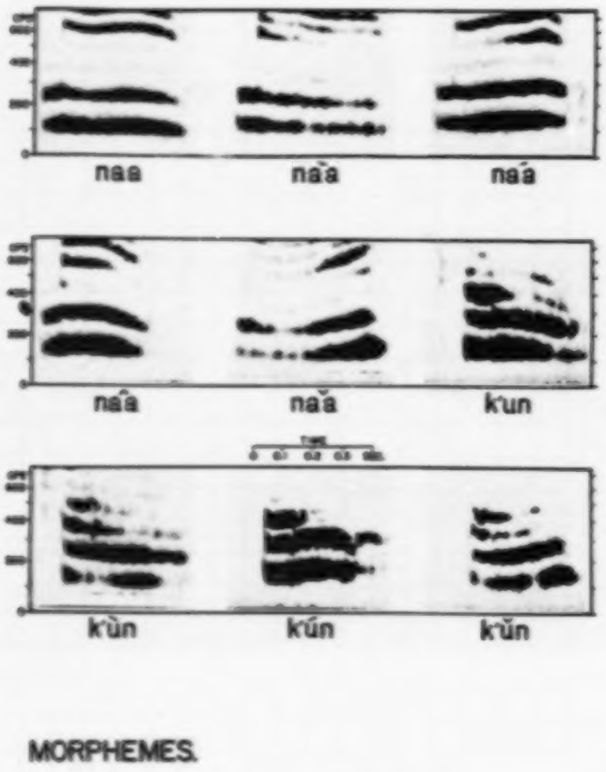


Fig. 23

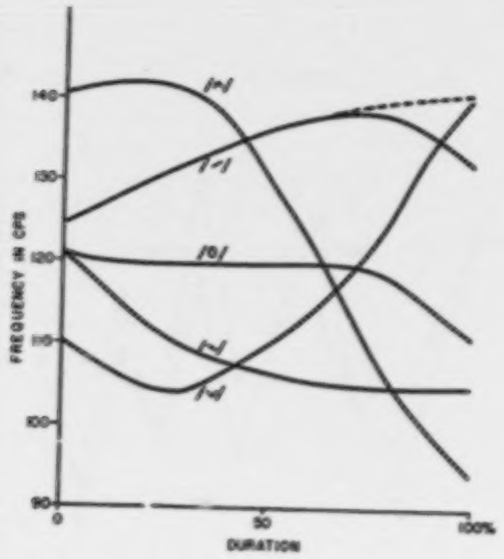


Fig. 24

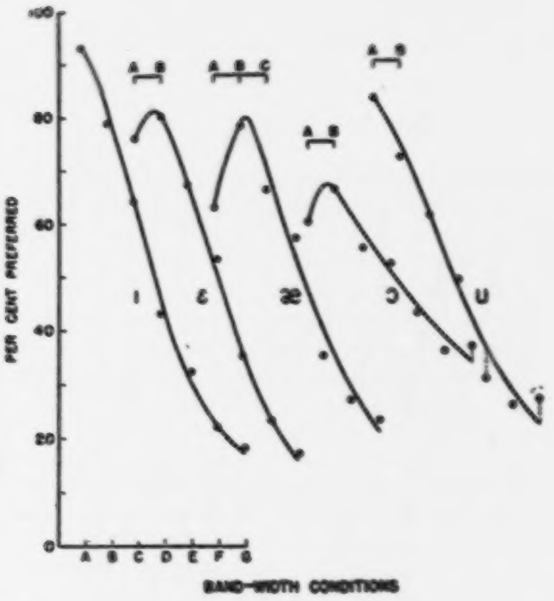


Fig. 25

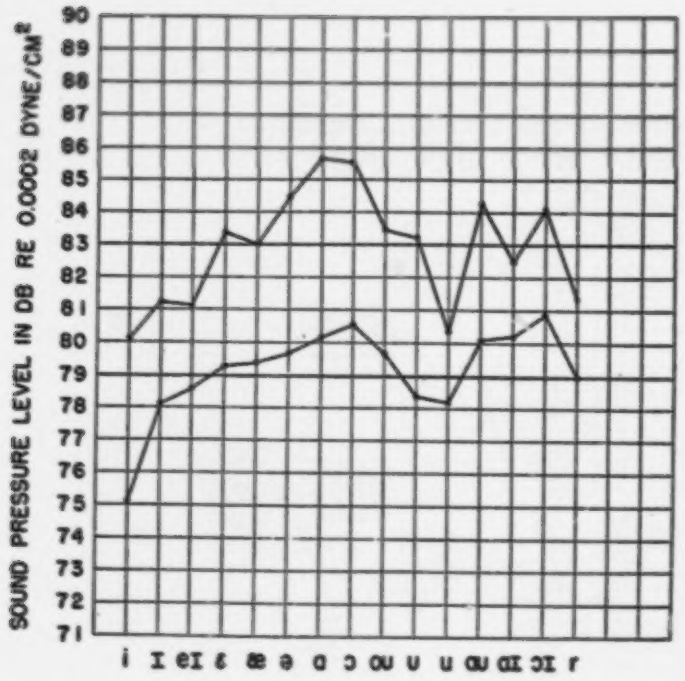


Fig. 26

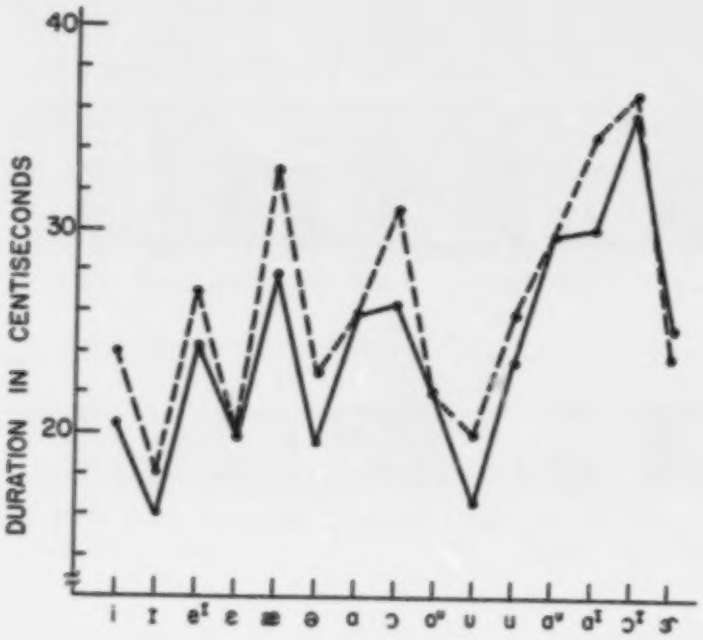


Fig. 27

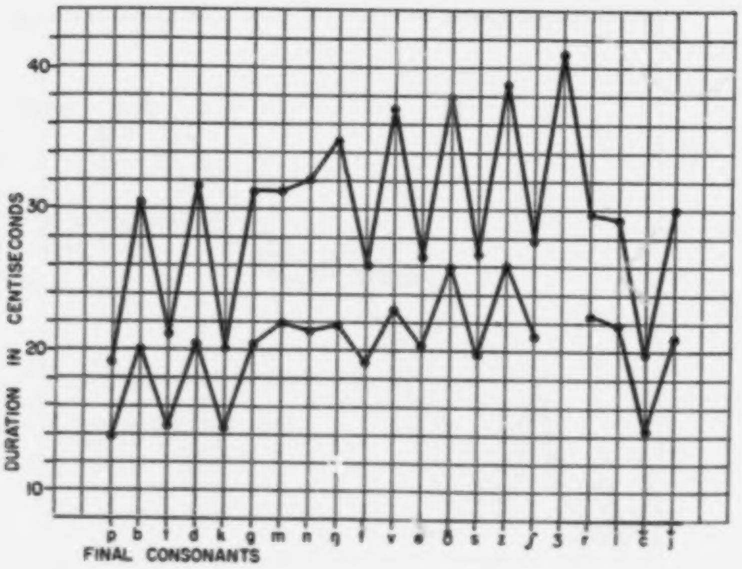


Fig. 28

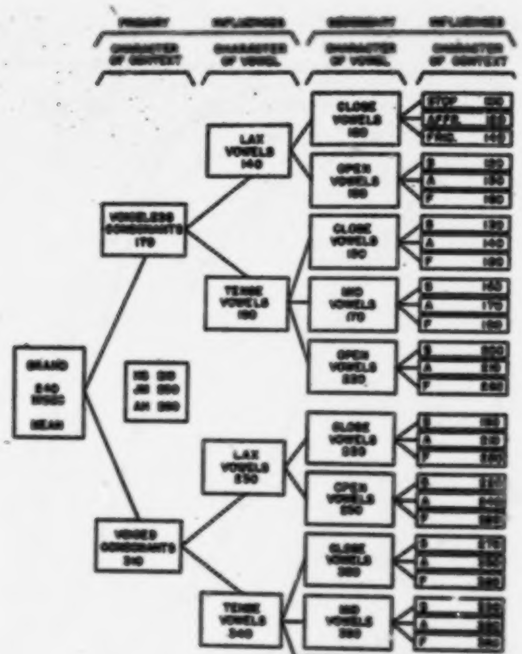

PRIMARY INFLUENCES
SECONDARY INFLUENCES
CHARACTER OF CONTEXT
CHARACTER OF VOWEL
CHARACTER OF VOWEL
CHARACTER OF CONTEXT
GRAND MEAN
VOICELESS CONSONANTS 170
VOICED CONSONANTS
LAX VOWELS 140
TENSE VOWELS
LAX VOWELS
TENSE VOWELS
CLOSE VOWELS
OPEN VOWELS
CLOSE VOWELS
MID VOWELS 170
OPEN VOWELS
CLOSE VOWELS
OPEN VOWELS
CLOSE VOWELS
MID VOWELS
OPEN VOWELS
STOP
AFFR.
FRIC.

Fig. 29

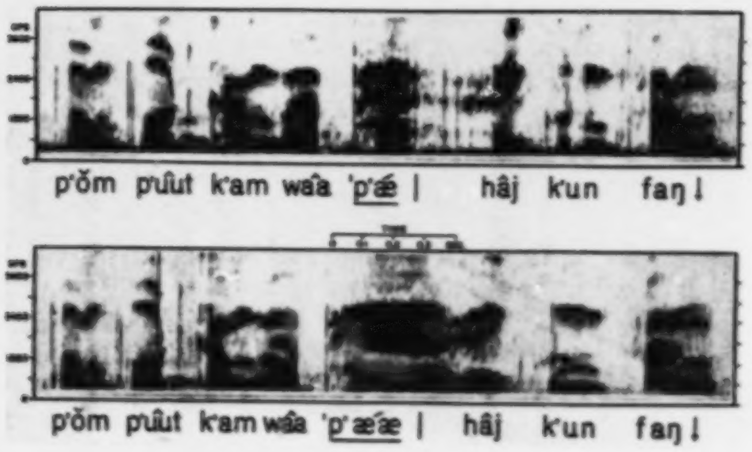


Fig. 30

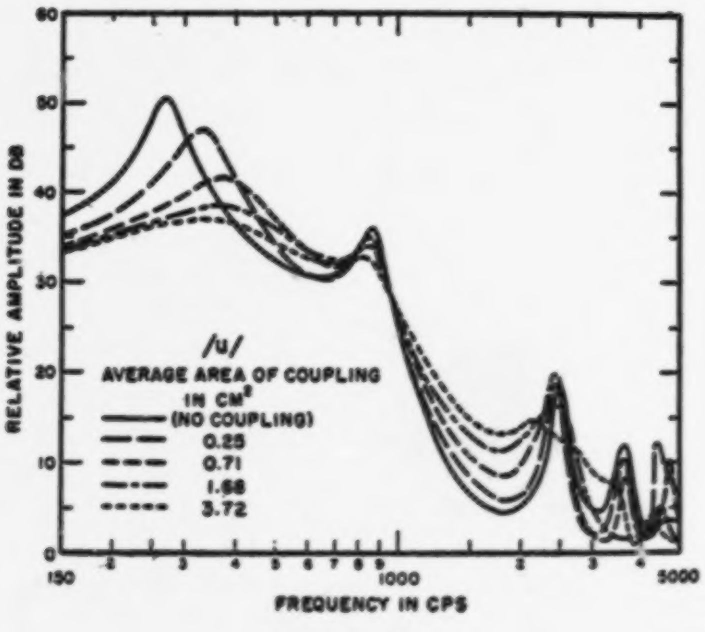


Fig. 31

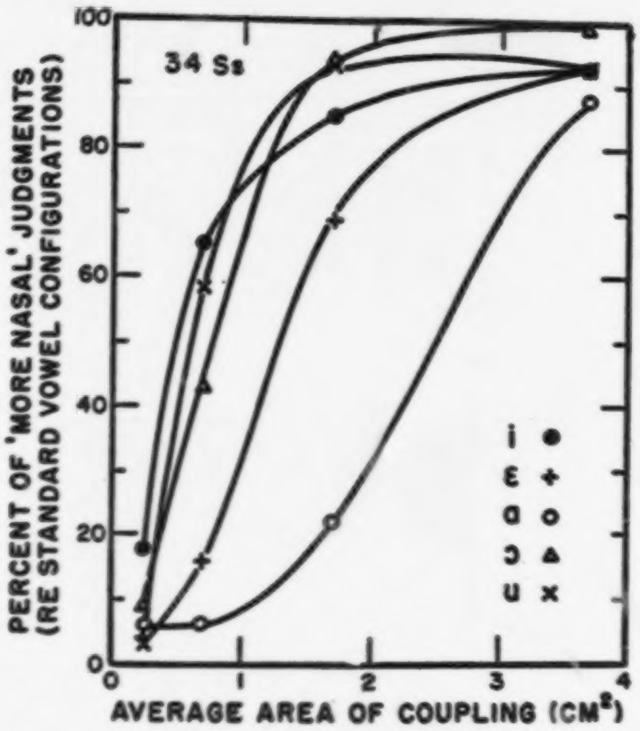


Fig. 32

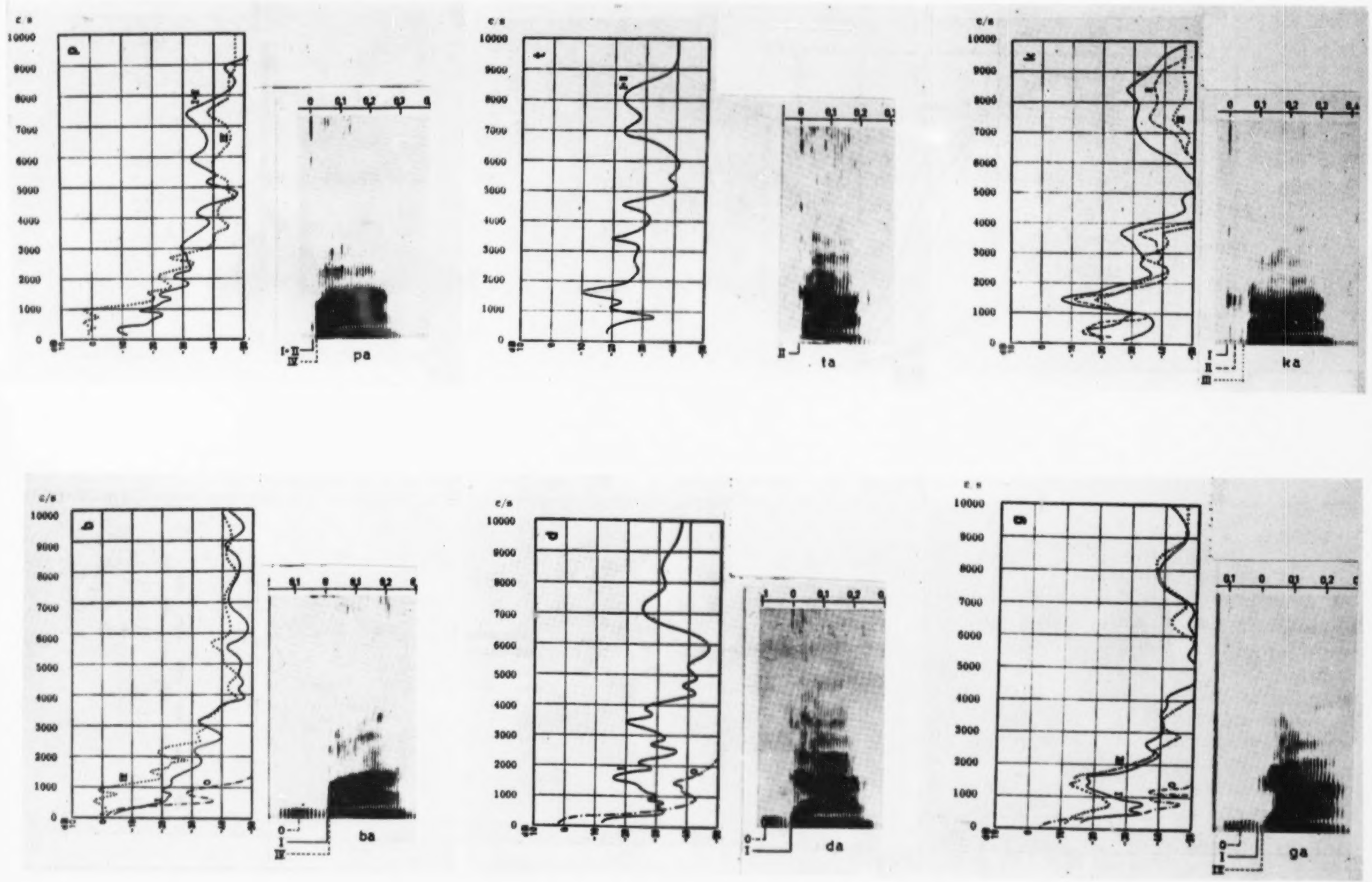

c/s
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
p
t
k
b
d
g
0
0.1
0.2
0.3
0.4
I
II
III
IV
O
pa
ta
ka
ba
da
ga

Fig. 33

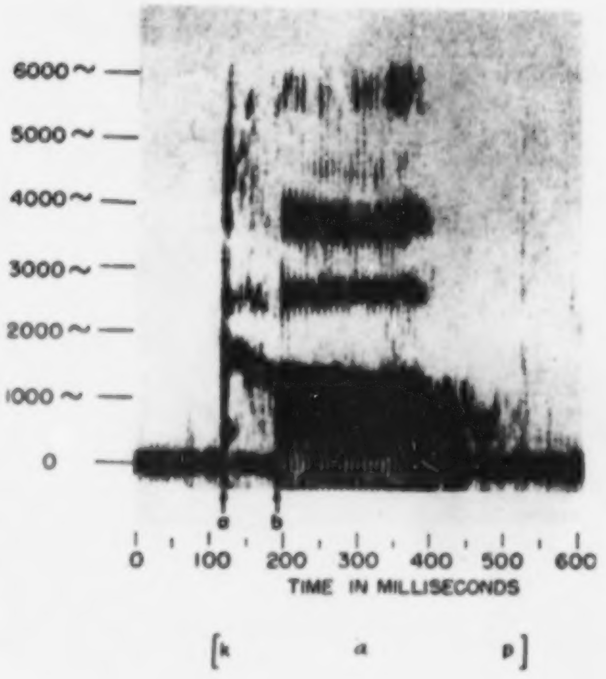


Fig. 34

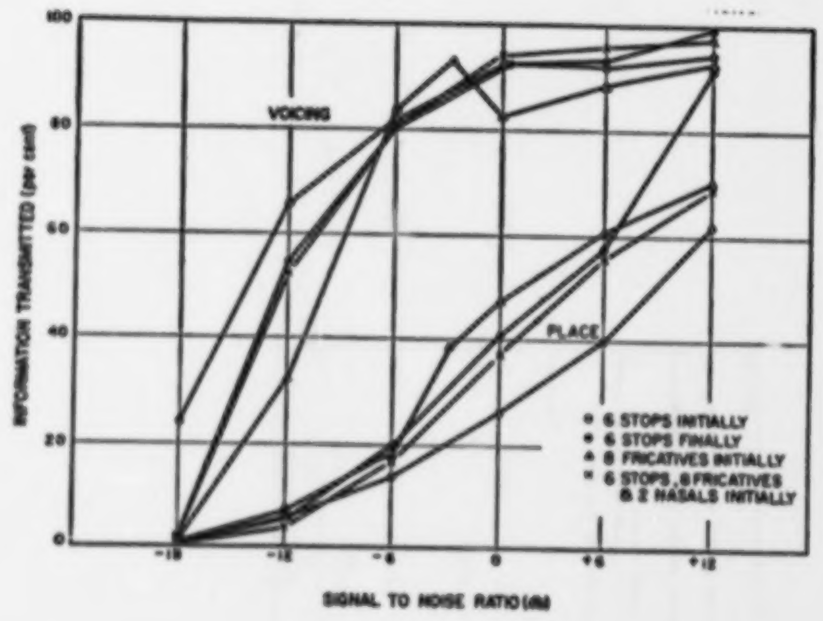


Fig. 35

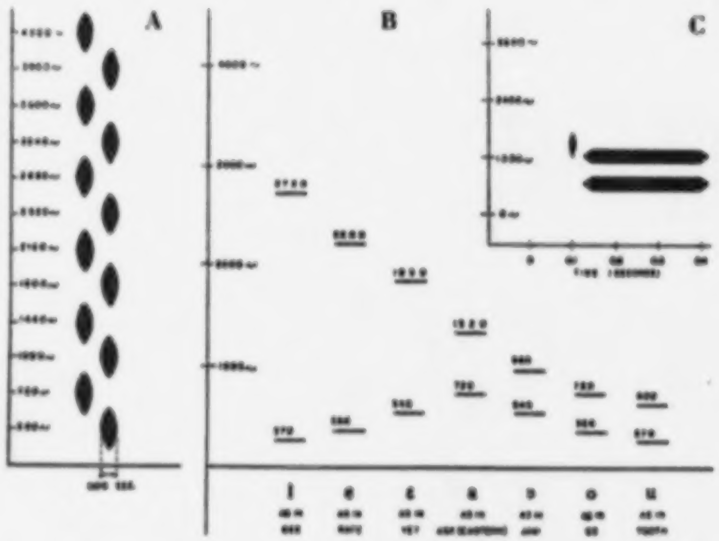

A
B
C
i
e
ɛ
a
ɔ
o
u

Fig. 36

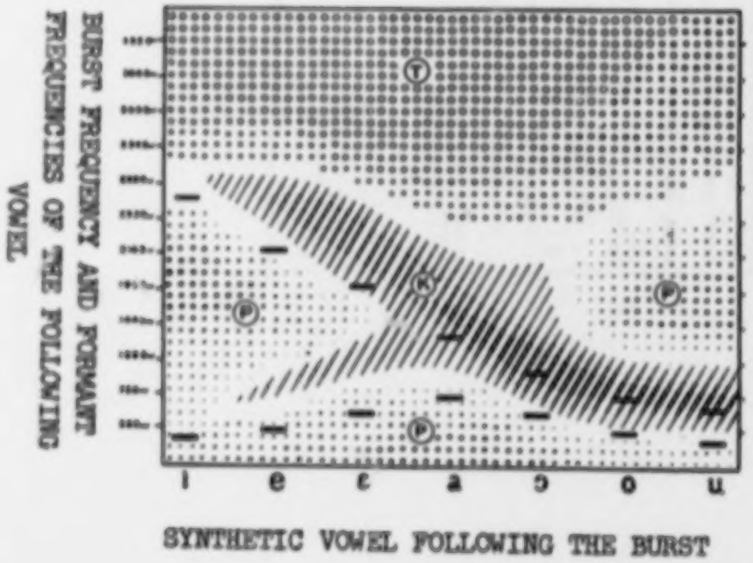


Fig. 37

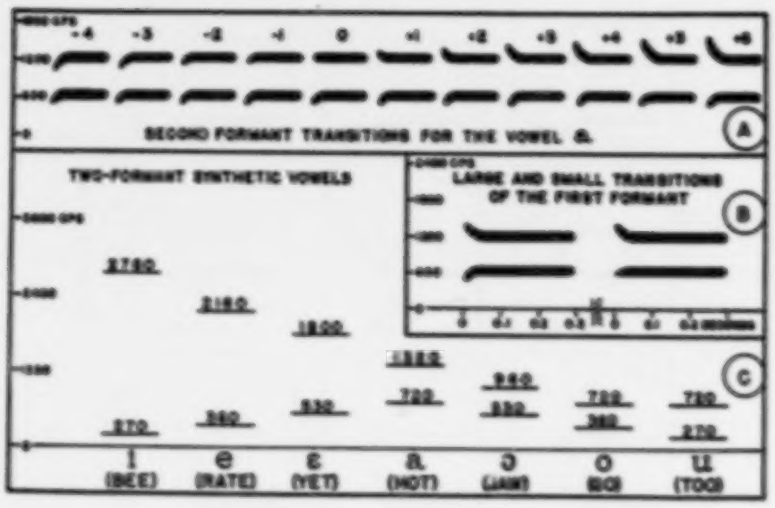

-4
-3
-2
-1
0
+1
+2
+3
+4
+5
+6
SECOND FORMANT TRANSITIONS FOR THE VOWEL a
A
TWO-FORMANT SYNTHETIC VOWELS
LARGE AND SMALL TRANSITIONS OF THE FIRST FORMANT
B
0 0.1 0.2 0.3 0 0.1 0.2 SECONDS
C
2760
270
2160
360
1800
530
1320
720
960
530
720
360
720
270
i (BEE)
e (RATE)
ɛ (YET)
a (HOT)
ɔ (JAW)
o (GO)
u (TOO)

Fig. 38

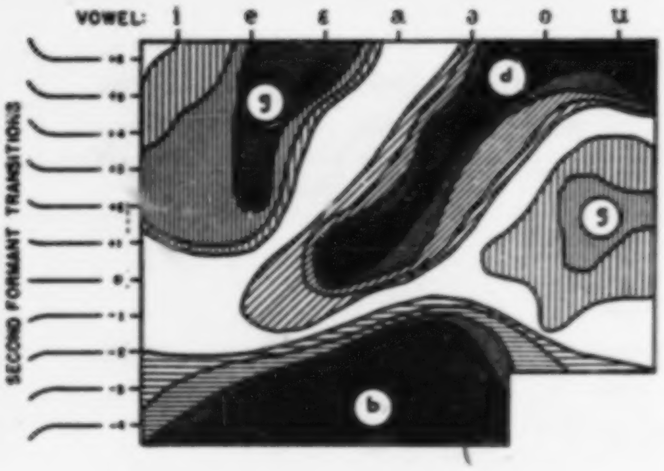


Fig. 39

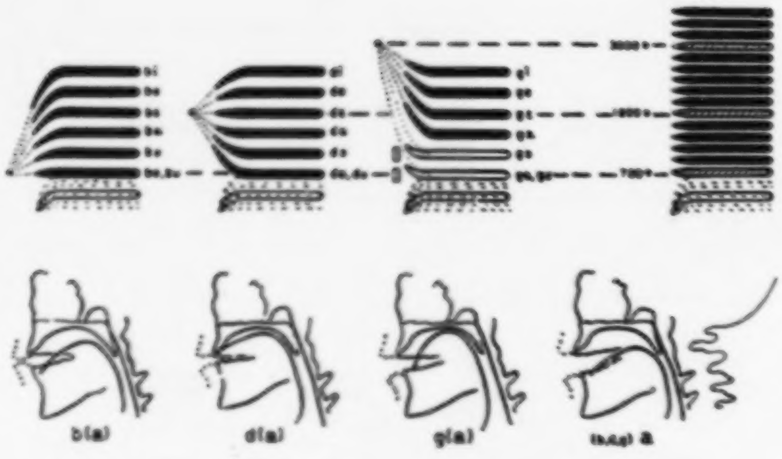

3000
1800
720
b(a)
d(a)
g(a)
(b,d,g) a

Fig. 40

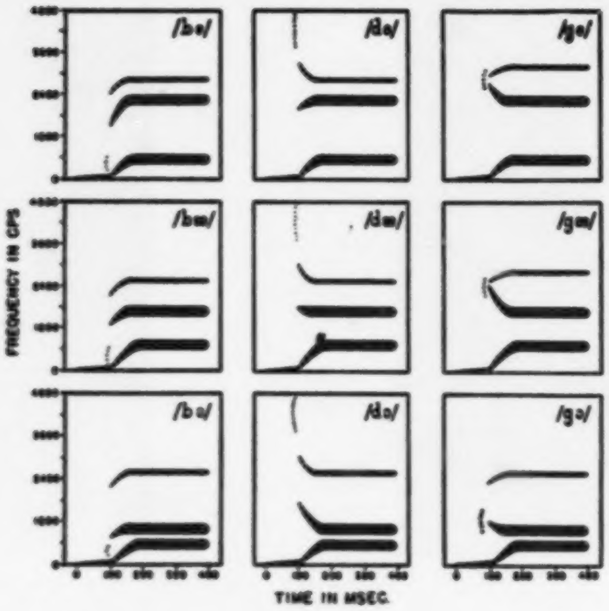


Fig. 41

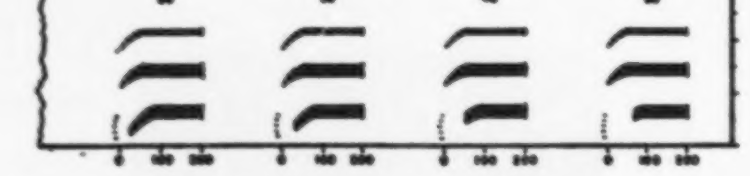


Fig. 42

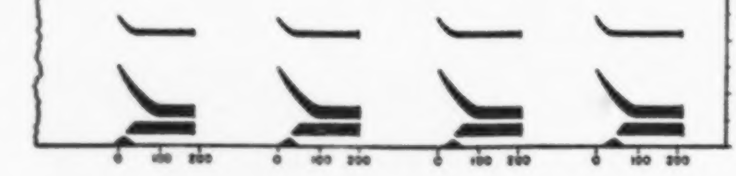


Fig. 43

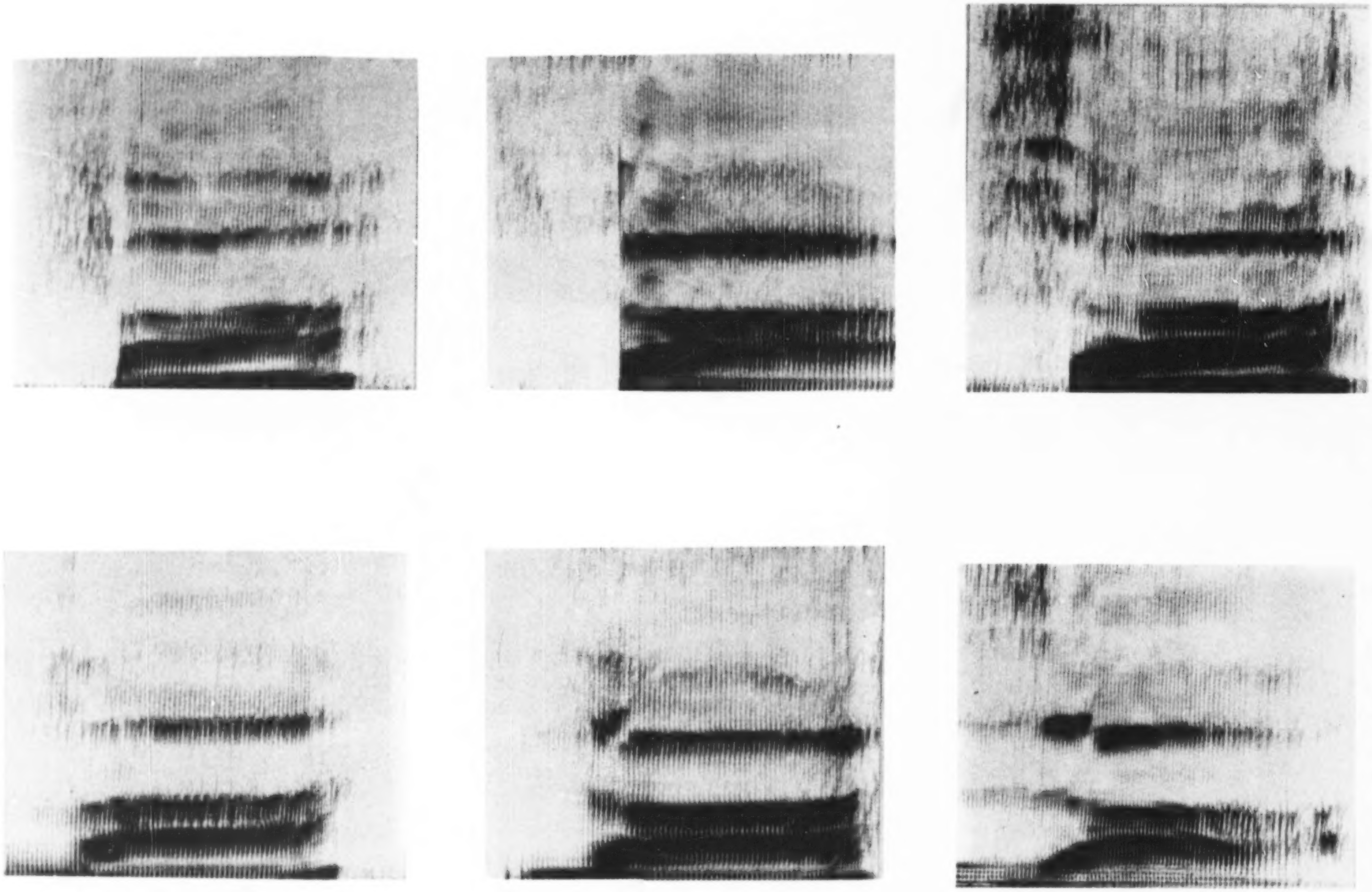


Fig. 44

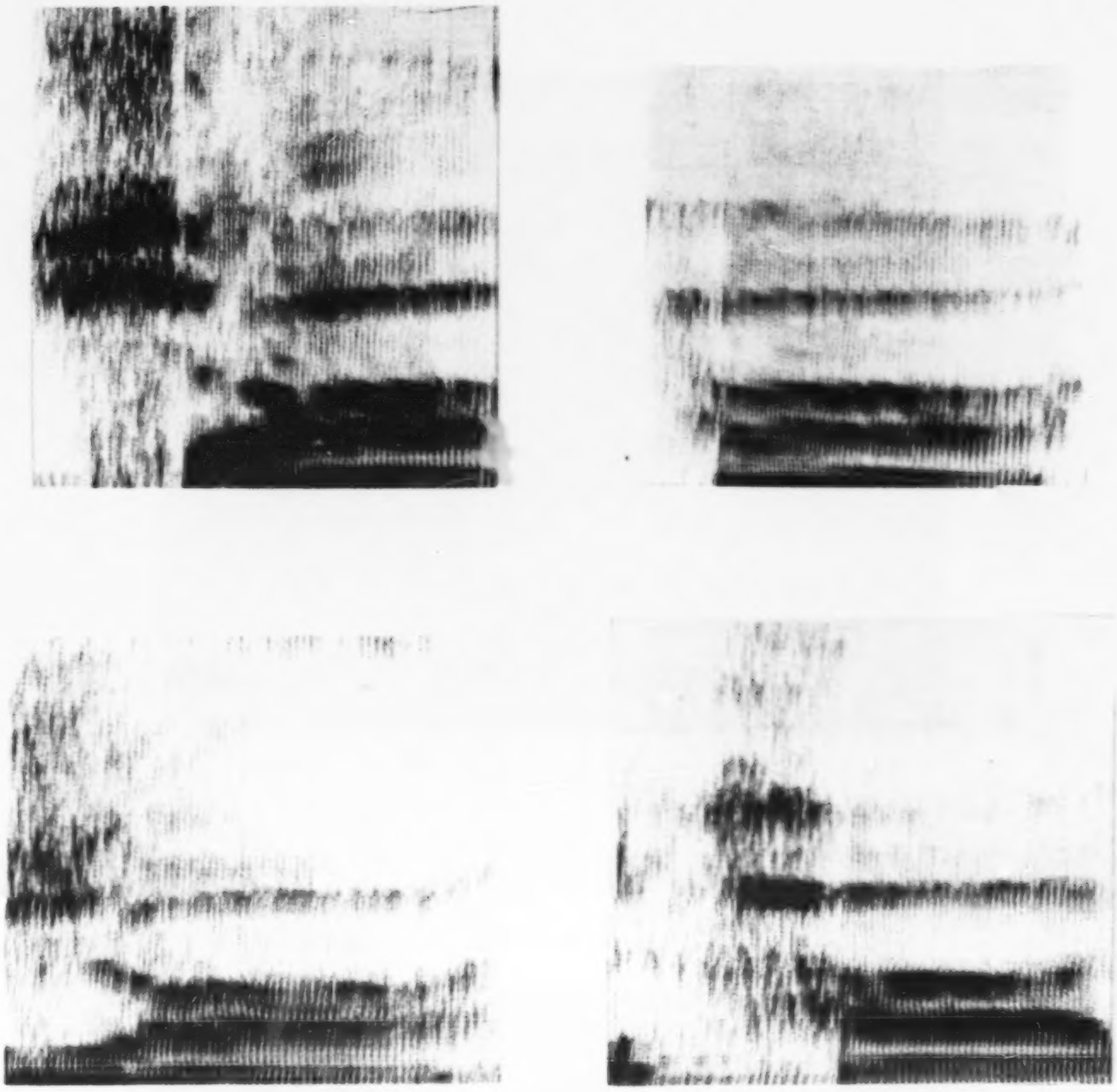
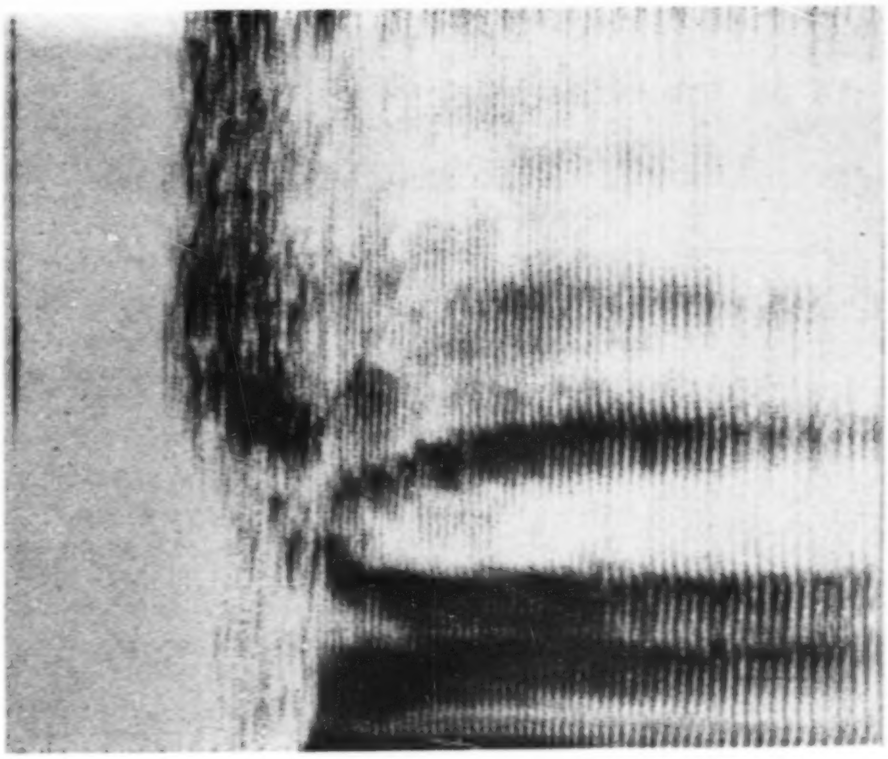


Fig. 45



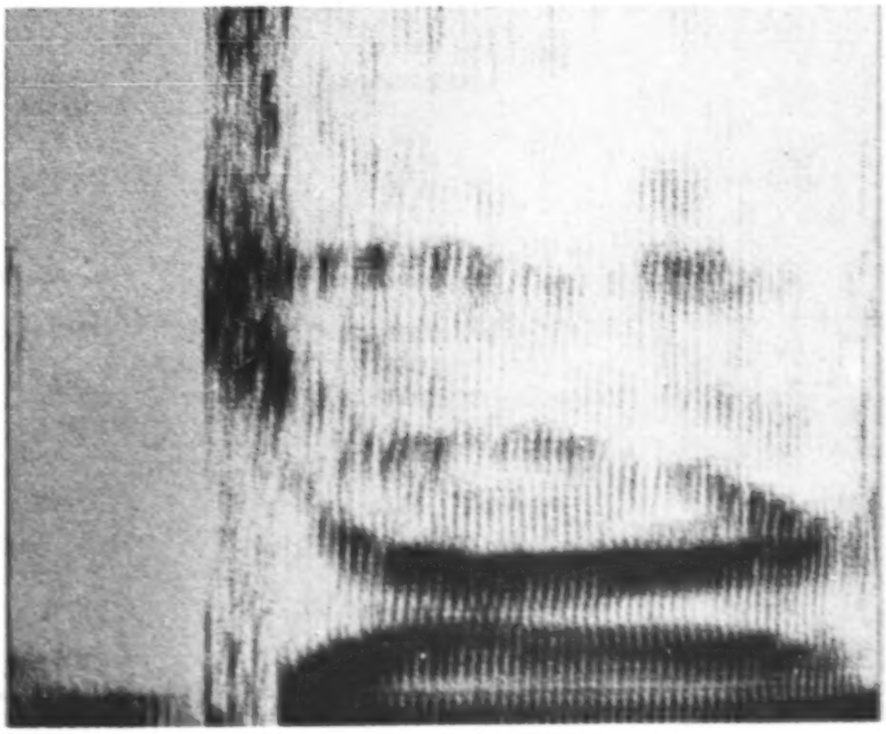


Fig. 46

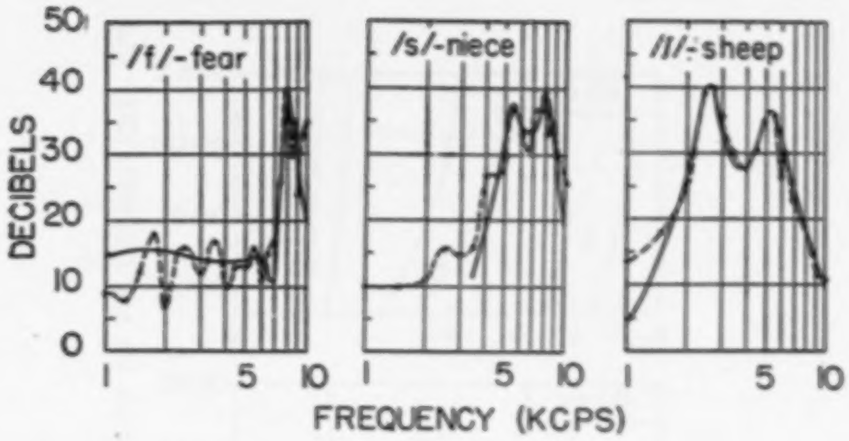


Fig. 47

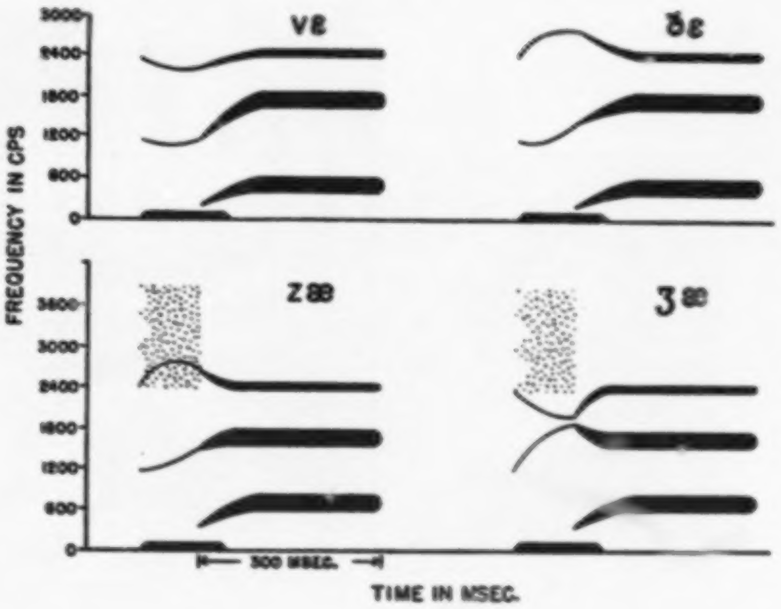

vɛ
ðɛ
zæ
ʒæ
FREQUENCY IN CPS
3000
2400
1800
1200
600
0
3600
300 MSEC.
TIME IN MSEC.

Fig. 48

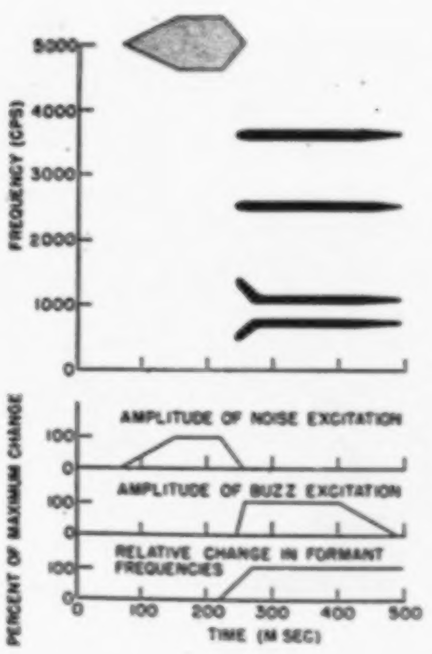

5000
4000
3000
2000
1000
0
FREQUENCY (CPS)
AMPLITUDE OF NOISE EXCITATION
AMPLITUDE OF BUZZ EXCITATION
RELATIVE CHANGE IN FORMANT FREQUENCIES
PERCENT OF MAXIMUM CHANGE
100
0
100
200
300
400
500
TIME (M SEC)

Fig. 49

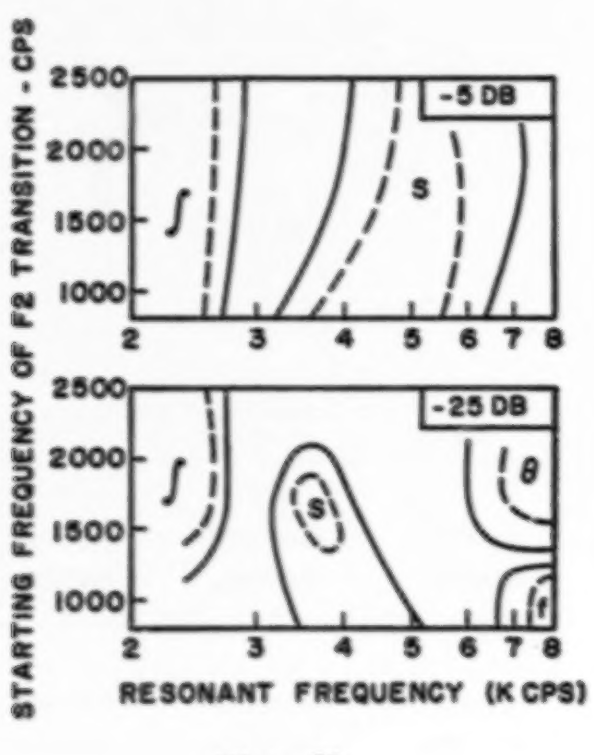

STARTING FREQUENCY OF F2 TRANSITION - CPS
-5 DB
-25 DB
ʃ
s
θ
f
2500
2000
1500
1000
2
3
4
5
6
7
8
RESONANT FREQUENCY (K CPS)

Fig. 50

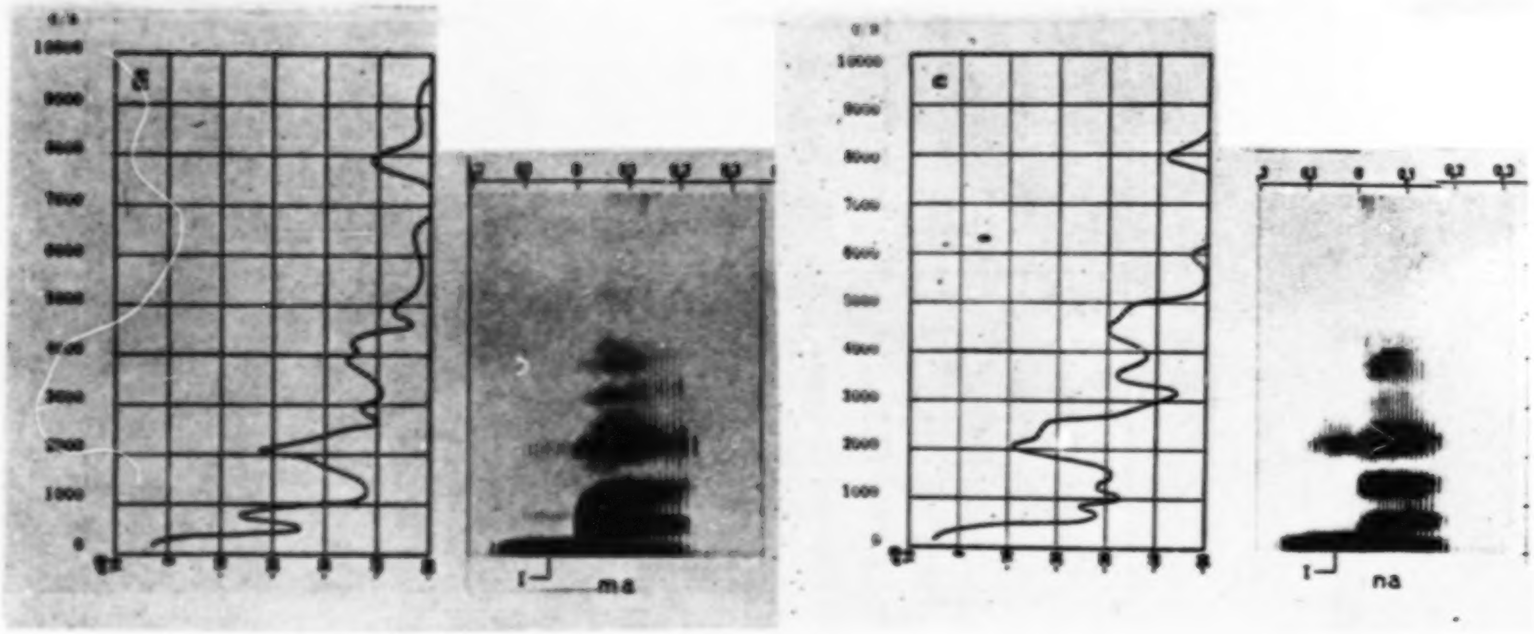

ʃ
c
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
ms
na

Fig. 51

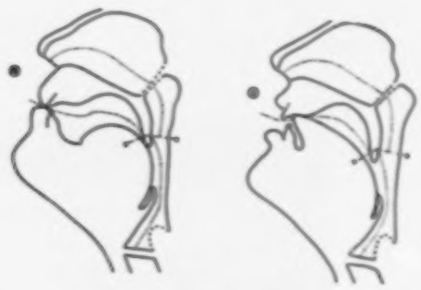


Fig. 52

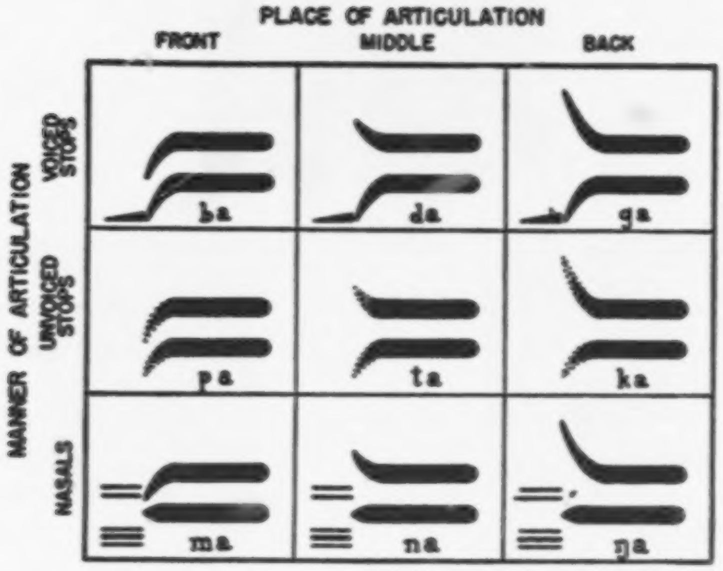

PLACE OF ARTICULATION
FRONT
MIDDLE
BACK
MANNER OF ARTICULATION
VOICED STOPS
UNVOICED STOPS
NASALS
ba
da
ga
pa
ta
ka
ma
na
ŋa

Fig. 53

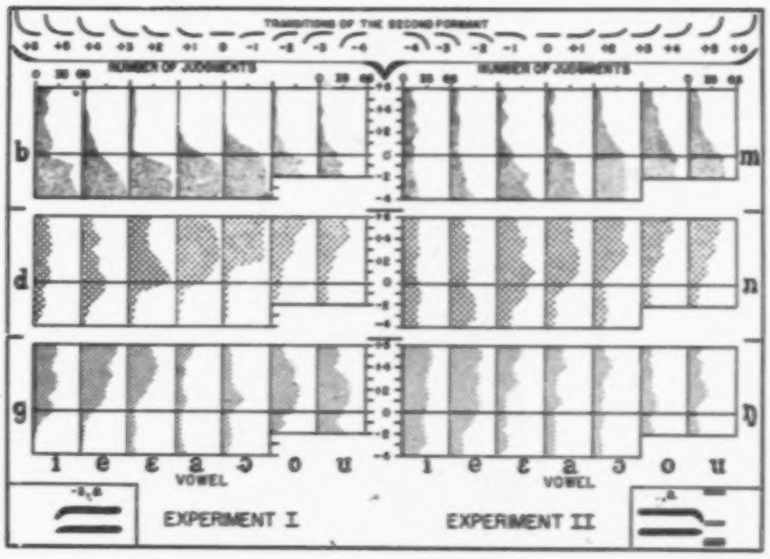


Fig. 54

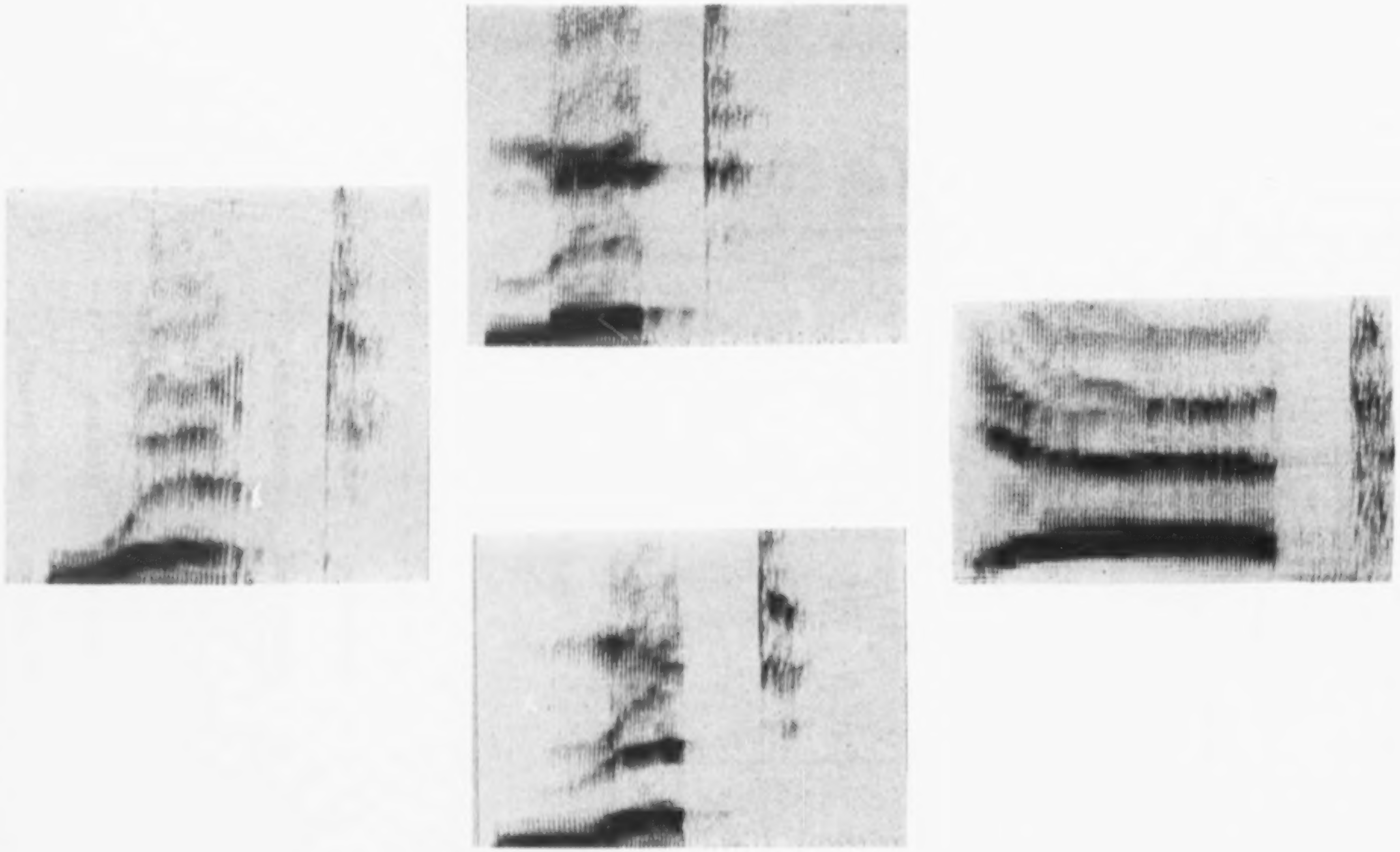


Fig. 55

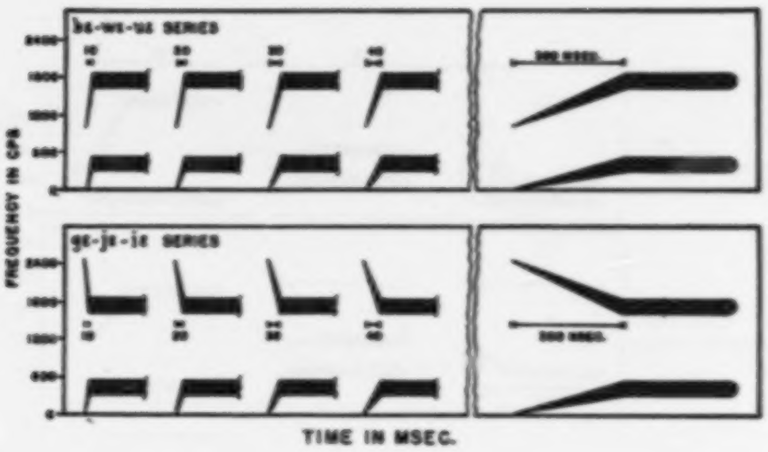

bɛ-wɛ-uɛ SERIES
gɛ-jɛ-iɛ SERIES
2400
1800
1200
600
0
10
20
30
40
300 MSEC.
FREQUENCY IN CPS
TIME IN MSEC.

Fig. 56

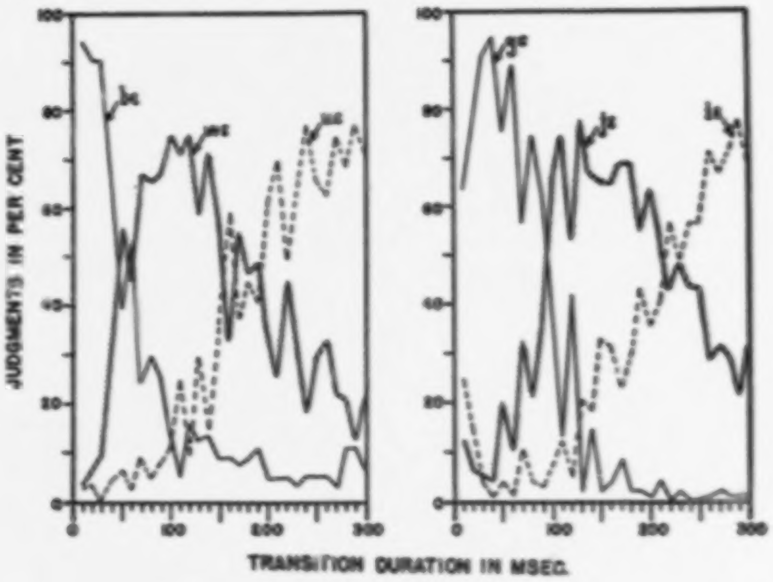


Fig. 57

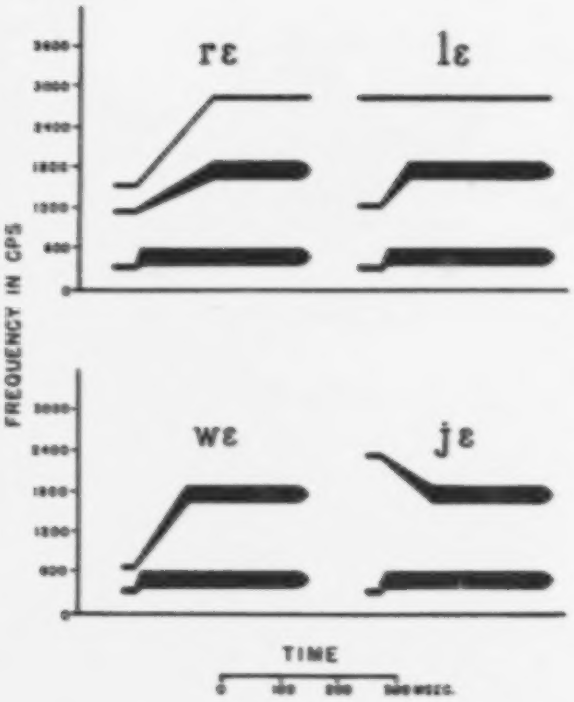

rɛ
lɛ
wɛ
jɛ
3600
3000
2400
1800
1200
600
0
FREQUENCY IN CPS
TIME
0
100
200
300 MSEC.

Fig. 58

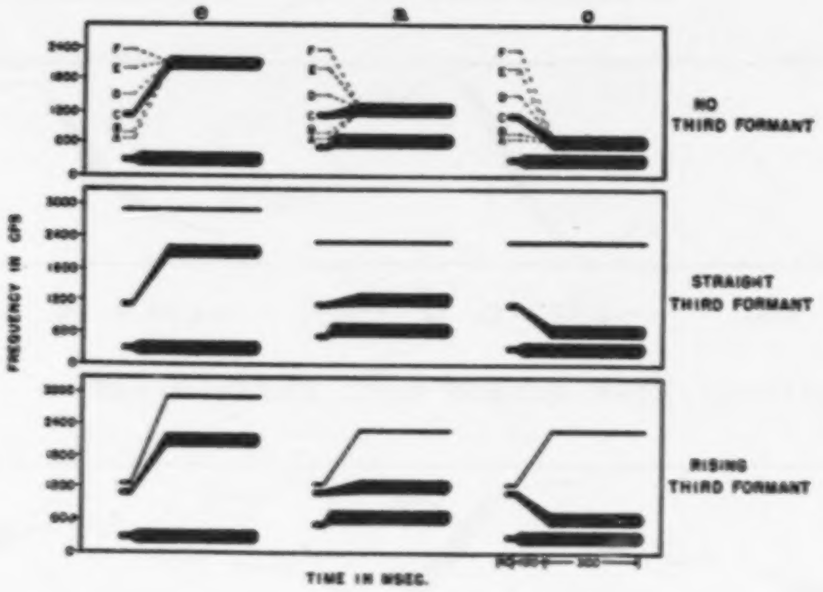

e
a
o
NO THIRD FORMANT
STRAIGHT THIRD FORMANT
RISING THIRD FORMANT
FREQUENCY IN CPS
TIME IN MSEC.

Fig. 59

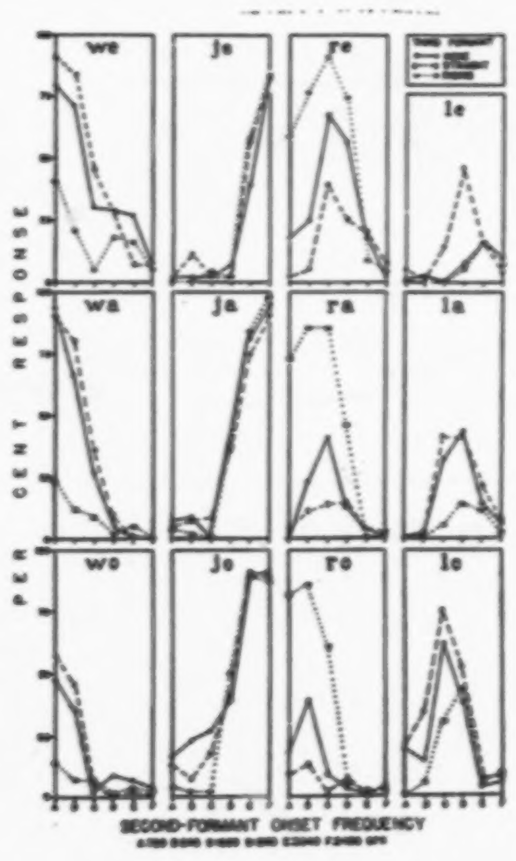


Fig. 60

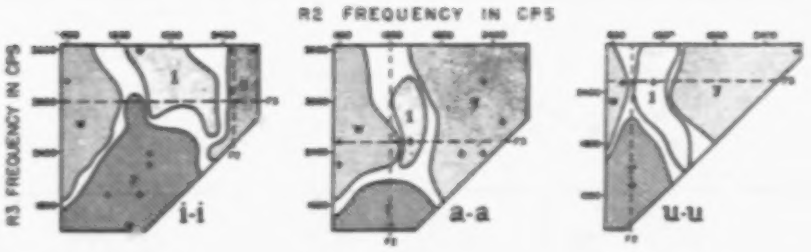


Fig. 61

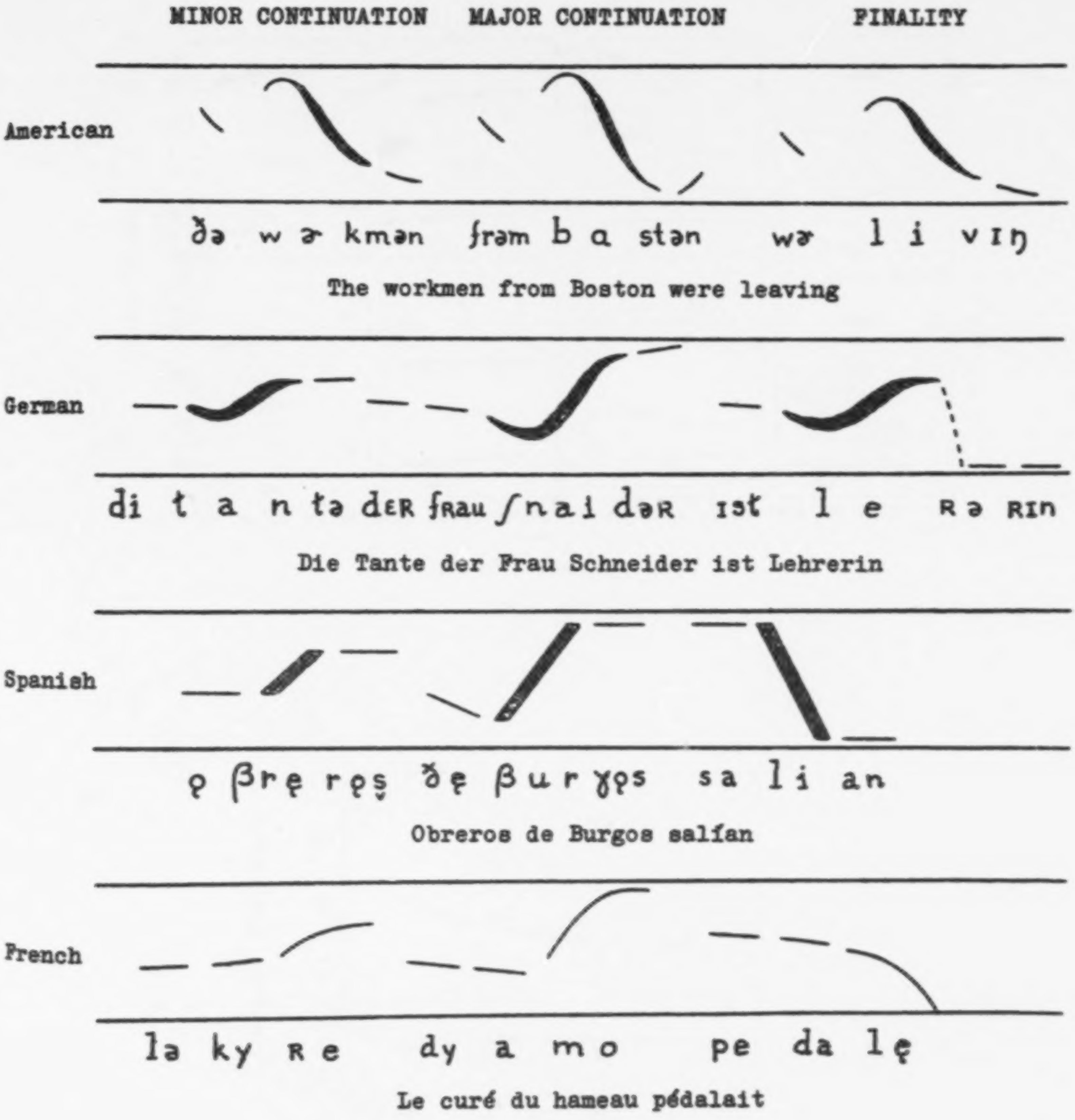

MINOR CONTINUATION
MAJOR CONTINUATION
FINALITY
American
ðə w ɚ kmən ʃrəm b ɑ stən wɚ l i vɪŋ
The workmen from Boston were leaving
German
di t a n tə dɛʀ ʃʀau ʃnai dəʀ ɪst l e ʀə ʀɪn
Die Tante der Frau Schneider ist Lehrerin
Spanish
ǫ βrę rǫs̬ ðę βurɣǫs sa li an
Obreros de Burgos salían
French
lə ky ʀe dy a mo pe da lę
Le curé du hameau pédalait

Fig. 62

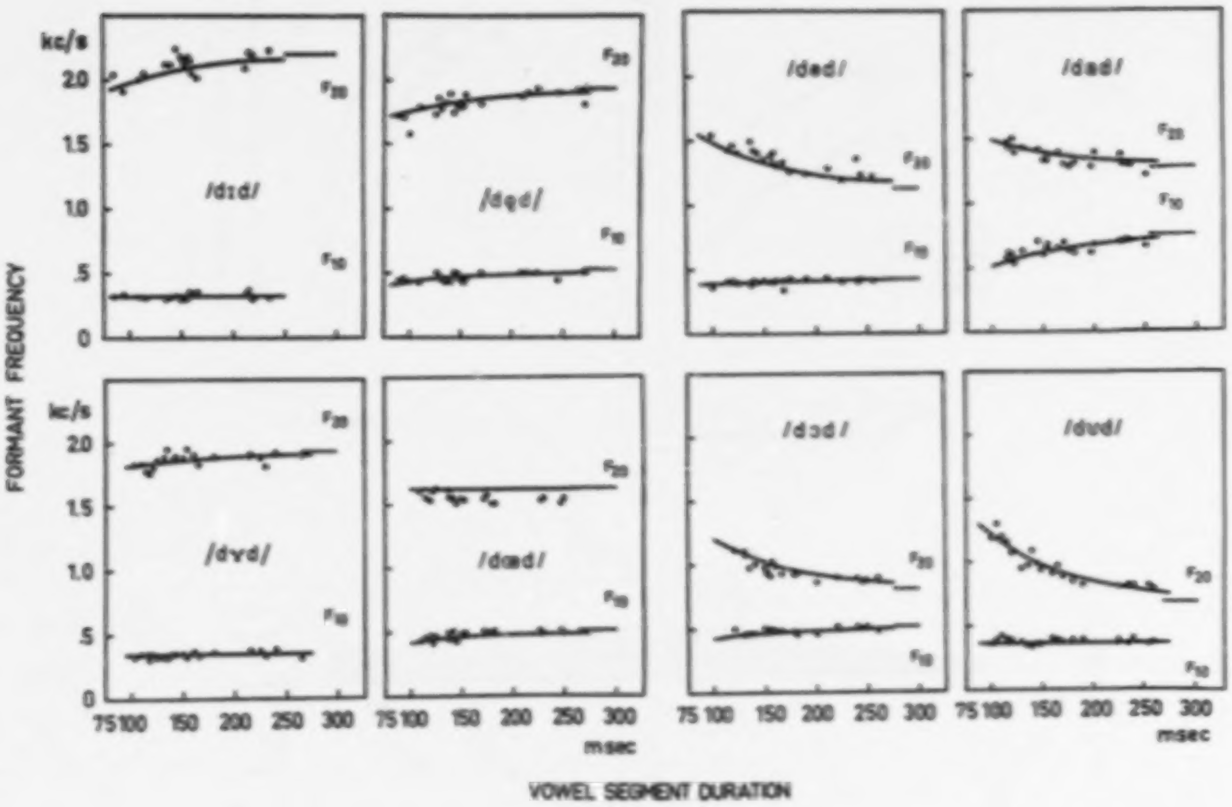

kc/s
2.0
1.5
1.0
.5
0
FORMANT FREQUENCY
/dɪd/
/dɛd/
/ded/
/dad/
/dʏd/
/dæd/
/dɔd/
/dud/
F_{20}
F_{10}
75
100
150
200
250
300
msec
VOWEL SEGMENT DURATION

Fig. 63

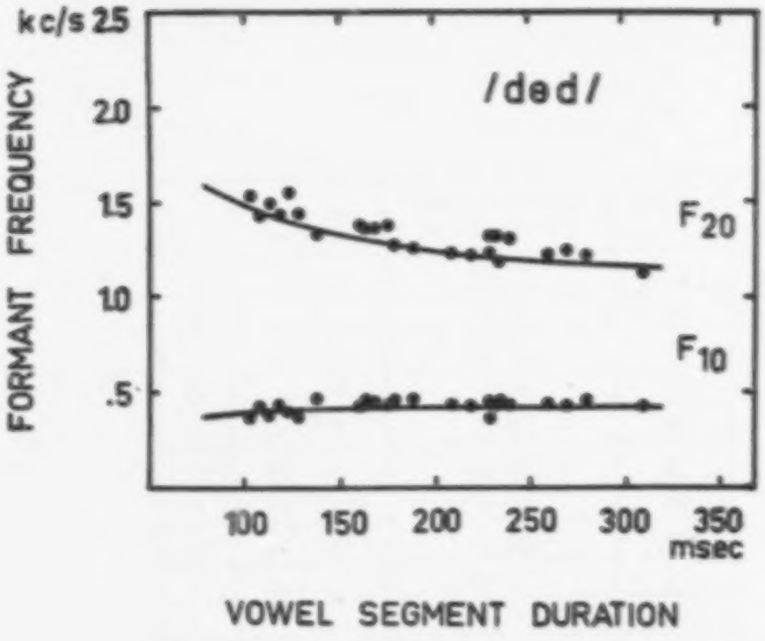

kc/s
2.5
2.0
1.5
1.0
.5
FORMANT FREQUENCY
/ded/
F_{20}
F_{10}
100
150
200
250
300
350
msec
VOWEL SEGMENT DURATION

Fig. 64

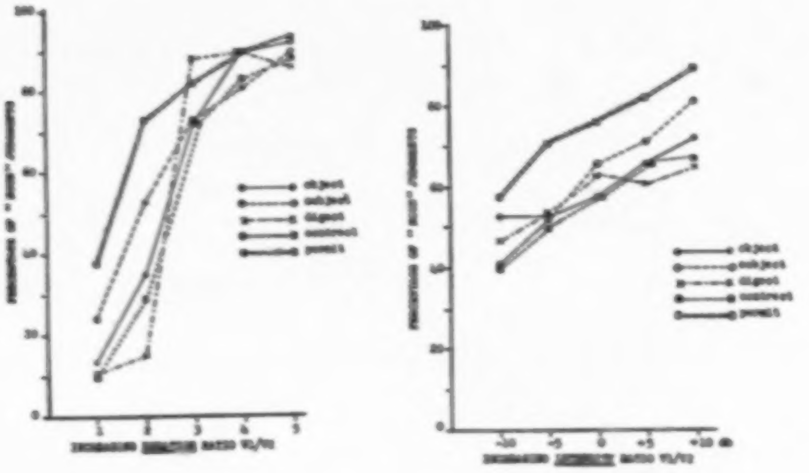


Fig. 65

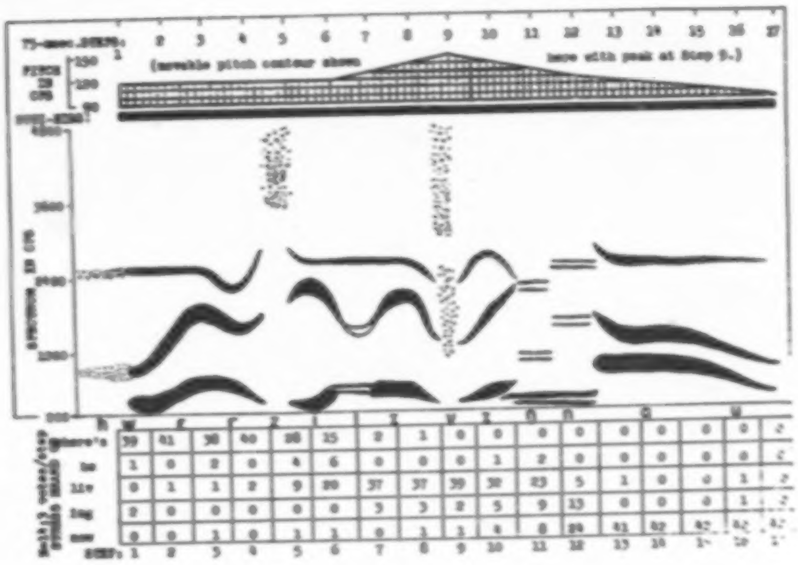
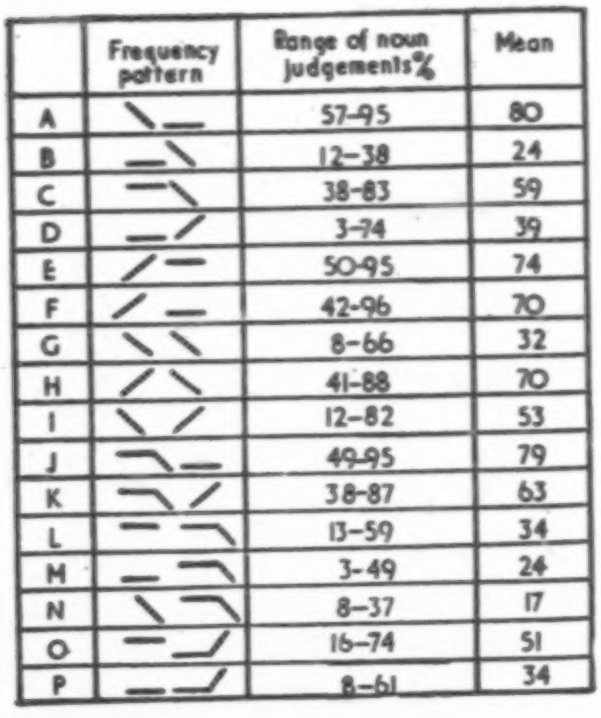

(movable pitch contour shown
here with peak at Step 9.)
PITCH IN CPS
150
120
90
FREQUENCY IN CPS
1000
2000
3000
there's
be
liv
log
now
STEP:

Fig. 66



	Frequency pattern	Range of noun judgements%	Mean
A	╲ —	57-95	80
B	— ╲	12–38	24
C	‾╲	38-83	59
D	— ╱	3-74	39
E	╱ ‾	50-95	74
F	╱ —	42-96	70
G	╲ ╲	8-66	32
H	╱ ╲	41-88	70
I	╲ ╱	12–82	53
J	‾╲ —	49-95	79
K	‾╲ ╱	38-87	63
L	‾ ‾╲	13–59	34
M	— ‾╲	3-49	24
N	╲ ‾╲	8–37	17
O	‾ _╱	16–74	51
P	— _╱	8–61	34

Fig. 67

MATCHING FUNCTIONS AND EQUAL-SENSATION CONTOURS FOR LOUDNESS

Strange Ross
The University of Michigan

Matching Functions and Equal-Sensation Contours for Loudness

Strange Ross

Behavior Analysis Laboratory

The University of Michigan

In the present study the "matching technique" is employed to investigate the input-output properties of the hearing mechanism with respect to the psychological attribute "loudness". The general question asked is: how does the loudness of a pure tone vary as a function of the frequency and the intensity of the tone? This particular formulation of the question, however, implies that "loudness" can be measured on an additive scale. For reasons which we have stated in detail elsewhere (Ross, 1964) we do not believe this to be possible at present. Therefore, we must rephrase our original question into two new questions which can be answered by the "matching technique": (1) Given two equally loud tones of different frequencies, what relationship must obtain between increments (or decrements) of the intensities of these separate tones in order to maintain the relation between them of equal loudness? and (2) For a change in the frequency (or intensity) of a pure tone of given loudness, how much compensatory increment or decrement in its intensity (frequency) is necessary to maintain the same loudness?

In an intensity-frequency plot, an equal-loudness contour represents a class of tones with the common property that all of these tones sound equally loud to a listener. Further, given two equal-loudness contours, the upper contour represents a class of tones with the common property that any one of these tones sounds louder to a listener than any tone belonging to the class of tones represented by the lower contour. If these two characterizations are valid, a complete set of equal-loudness contours determines the conversion of the physical dimensions "intensity" and "frequency" of pure tones into the psychological dimension "loudness". This psychological dimension is described as "intensive" because its elements (pure tones) can be *ordered* consistently (not excluding that two or more tones occupy the same place in this order). Such an order forms the basis for an *ordinal measurement* of the dimension in question. (For an extensive discussion of the logic of this kind of measurement, see Ross (1964).) A necessary and sufficient condition for the validity of the above-mentioned characterizations is that the equal-loudness relation be both symmetric and transitive.

Previous studies (Kingsbury, 1927; Fletcher & Munson, 1933; Churcher & King, 1937; Zwicker & Feldtkeller, 1955; Robinson & Dadson, 1956) have established equal-loudness relations between a standard tone of fixed frequency and comparison tones of different frequencies. On the basis of these data, sets of equal-loudness contours have been constructed. It has never been established, however, that the equal-loudness relations represented by these contours are both symmetric and transitive. In this study we intend to test explicitly whether or not the equal-loudness relation in fact has these formal properties.

When attempting to collect equal-loudness data, two general avenues of approach are possible. One of these consists in collecting data at each experimental setting "along an equal-loudness contour", i.e., varying the frequency of the variable between each adjustment to a fixed standard. The other approach is to collect data at each experimental setting "along a loudness matching function", i.e., keeping the frequencies of the standard and the comparison tones fixed, and varying the intensity of the standard after each adjustment of the comparison tone. Under the assumption that equal-loudness contours have a rather simple form when plotted in sensation level units, the former method has the disadvantage that fitting a curve to a direct plot of the data will be rather difficult. This is due to the fact that, when the data are plotted with either sound pressure level, or voltage across the sound transducer, as the ordinate, the individual absolute threshold curve (which may show pronounced irregularities above 1kcs) as well as the transmission characteristic of the sound transducer employed will make for a more or less complicated shape of the equal-loudness contours. The latter method, in which the frequencies of the standard and the comparison tones are kept constant, has the advantage that the obtained data distributions would be independent of both the individual absolute threshold curve and the transmission characteristic of the sound transducer. Also, this latter method yields directly pertinent information as to the shape of the loudness matching function.

Apparatus

The functions required of the experimental set-up were the following: 1) To present alternately two tones of different frequencies with no silent interval between the tones, and with no audible switching transients. 2) To allow the subject to adjust continuously the intensity of one of the two alternating tones (the comparison tone). 3) Following a signal from the subject, indicating the completion of a loudness match, to make a permanent record of the intensities of the two tones. 4) After each adjustment, to increase the intensity of the standard tone by a predetermined amount. 5) To make a corresponding increase in the level of the comparison tone, in order to maintain an appropriate intensity range of this tone.

6) After each completed adjustment, to introduce various amounts of attenuation, randomly programmed, of the comparison tone in order to eliminate possible cues from the position of the subject's potentiometer. 7) To present the two tones, at all intensity levels, with a low degree of distortion. 8) To maintain, at all intensity levels, a high signal-to-noise ratio for the two signals. 9) To ensure a high degree of reproducibility of the sound pressure levels presented to the subject. 10) To permit an absolute determination of the sound pressure levels presented, as measured at the subject's eardrum.

These objectives were accomplished in the following way. The standard tone was generated by an audio oscillator (General Radio 1304-B). The signal was amplified (General Radio 1206-B) and then fed through two attenuators to one channel of an electronic switch (Grason-Stadler 829D), set at a switching rate of 2.5 secs/cycle. One of the attenuators mentioned served to adjust the general signal level in order to obtain the highest signal-to-noise ratio possible, while the other attenuator was specially constructed to permit the automatic increase in intensity of the standard tone by a predetermined multiple of .5 db. This "stepping" attenuator consisted of two 10-section ladder attenuators in steps of .5 db and 5 db, respectively. Each ladder was terminated by a cathode follower in order to obtain the high terminating impedance required for this type of attenuator, and the 10 sections of each ladder were connected to the terminals of one of two sections of an electro-mechanical counter with electrical read-out (Neuron Counter 7005, Data Instruments). A number of operant conditioning devices (Grason-Stadler) generated, following the subject's signal, the number of pulses required to raise the intensity of the standard tone by a predetermined amount (in multiples of .5 db).

The comparison tone was generated by a second oscillator (Bruel & Kjaer 1014) which fed the signal through a "sone" potentiometer, controlled by the subject, and three attenuators to the other channel of the electronic switch. One of the three attenuators mentioned was a fixed attenuator which served to control the signal-to-noise ratio. The second was a "stepping" attenuator identical to the one described above. And the third attenuator was a relay-operated attenuator with four fixed values of 0, 2.5, 5.0, and 7.5 db; a stepping switch (Grason-Stadler E3126B) was used to select--according to a random schedule--one of the four possible attenuations after the completion of each adjustment.

In order to keep the distortion (arising from the cathode followers of the "stepping" attenuators and from the electronic switch) at a minimum, each output of the electronic switch was passed through a band-pass filter (Krohn-Hite 310-ABR) set at ±15% of the respective frequencies, and the outputs of the two filters were combined in a matching network. The output of this network was amplified (General Radio 1304-B, output section), and finally fed through an attenuator (General Radio 1450-TBR). The distortion of the final output was always below 1%.

Depression of the subject's key initiated the following operations: 1) The pen of a graphic level recorder (Bruel & Kjaer 2305) was momentarily lowered to make a dot on the recording paper, indicating the momentary voltage level at the output of the "stepping" attenuator; (through the calibration of the system, this voltage level was converted to the corresponding voltage level over the telephone). 2) The two "stepping" attenuators were advanced either 2.5db or 5.0db, depending on the experiment. 3) The recording paper was advanced by an amount proportional to the increment of the intensity of the standard tone. 4) The "random" attenuator was adjusted to its next value in a random series.

The signals from the final attenuator were applied to the subject's ear by means of a small electro-dynamical telephone (Beyer DT-508), mounted in an ear insert moulded to fit the external meatus of the individual subject. The ear insert contained a canal for the conduction of the sound, a steel tube through which a probe microphone could be inserted, and a narrow steel tube for equalizing the static pressure of the air on both sides of the ear inset. When a probe microphone was not inserted, a dummy with the same acoustical properties took its place. This technique for applying the stimulus (which has been described in greater detail elsewhere [Ross, 1963]) affords the following advantages: 1) Due to the individual moulding of each ear insert, the position of the sound-transducing elements can be closely replicated; this, in turn, means that from one experimental session to the next, the same sound pressure levels are always produced at the eardrum for given stimulus parameters. 2) These sound pressure levels can be measured, thereby providing a conversion from the electric parameters of the signals to absolute acoustic measures. 3) Due to the tight acoustic coupling between the telephone membrane and the enclosed small volume of air (approximately .2cc), the distortion of the acoustic signal is kept very low; for all frequencies employed, and for sound pressure levels up to 120 db SPL, the distortion due to the telephone was found to be less than 1%.

Procedure

At the outset of each experiment, four threshold determinations were made for each of the two frequencies involved in the series of matches to follow. The test tone was presented through the comparison channel, and the standard channel was disconnected. The test tone thus was on for

1.25 secs., off for 1.25 secs., etc. The subject was requested to adjust the tone to "just not audible" or to "just audible", whatever he preferred, and to press his key when the adjustment was completed.

Following these threshold determinations, the intensity of the standard tone was adjusted to approximately 5-10 db below the subject's threshold as estimated from the four threshold adjustments, and the subject was requested to adjust the intensity of the comparison tone to the same loudness as that of the standard tone, and to indicate a completed adjustment by pressing his key. After each key press the intensity of the standard tone was automatically increased by a predetermined number of dbs. During the first few trials, the subject would not be able to hear the standard tone, and according to the instructions he would indicate this by turning down the intensity of the comparison tone all the way. In this way loudness matches from levels very close to the thresholds were secured. The subject continued to match the two tones until he came close to an uncomfortable loudness level.

In one series of experiments, the standard tone was increased by 2.5 db after each adjustment. In this case, a complete run lasted approximately one hour, and the subjects were able to make three complete runs a day. In a later series, the increment was increased to 5 db after each adjustment. In this series subjects were able to make six complete runs a day. After each complete run the subjects rested for about 15 minutes.

Thirteen frequencies extending from 20 cps to 5000 cps, and spaced 1/5 decade apart, were employed; (the calibration procedure of the sound transducer, as explained elsewhere [Ross, 1963], prevented the inclusion of frequencies higher than 5000 cps).

To date, a total of 106 sets of data have been collected from one subject (female, 27 years, extensive previous experience in frequency discrimination). The frequency combinations were distributed according to Table 1. (The uneven number of repetitions for each frequency combination is due partly to the exploratory nature of the investigation, partly to the fact that when large discrepancies occurred between two sets of data, additional sets were collected.) The collection of data from this subject is yet to be completed.

The greater part of the data were collected in groups of three sets, thereby facilitating the tests for transitivity. In two of these the same tone at the same intensity values served as the standard tone, and in the third set the comparison tones of the first two sets were matched. Transitivity was checked by plotting against each other the intensity values of the comparison tones from the first two sets of data corresponding to the same intensity value of the standard tone, and comparing this plot to the data from the third set; for transitivity to hold, these two plots should coincide. In later experiments the frequencies of the standard and the comparison tones were reversed (some of these sets of data still remain to be collected) to provide a check on the symmetry of the "equal-loudness relation."

Results

As mentioned before, results from only one subject will be presented in this paper. Data from a second subject are currectly being collected, and it is planned to collect data from one or two additional subjects. Further, it should be stressed that the data analysis to be presented in the following is of a preliminary nature. At a later time, a more "objective" analysis will be carried out.

Symmetry and Transitivity

Figures 2 and 3 show the equal-loudness matches for two pairs of frequencies. One set of points (dots) represents loudness matches when the frequency labelling the ordinate served as the comparison tone; another set of points (circles) represents the loudness matches obtained when the frequency labelling the abscissa served as comparison tone. The third set of points (crosses) are computed from two other experiments in both of which a given intermediate frequency served as the standard tone. The dot and the circle within a circle represent the "matches" of the threshold intensities for the given experiments. It is clear that the loudness matches are practically independent of which tone is the comparison tone, that is, the "equal-loudness relation" as established through these experiments is symmetric. Further, the predicted set of points coincides closely with the obtained points, that is, we can conclude that the "equal-loudness relation" for these experiments is transitive. The matching functions presented for these two pairs of frequencies are representative of the results obtained for all other combinations of frequencies employed. We may therefore assume that the "equal-loudness relation" for any frequency combination, and at any loudness level, is both symmetric and transitive.

Conjoint Curve-Fitting

Accepting on the basis of the results stated above that the "equal-loudness relation" is both symmetric and transitive (excluding random errors), it should be possible to fit to each set of data a curve in such a way that the transitivity of the data is maintained. Let us illustrate this process by considering three sets of data collected for the three possible combinations of three different frequencies, F_1, F_2, and F_3. Because of the noise inherent in the data, a certain leeway is present in fitting a curve to each of the three sets of data. The assumed transitivity of the data implies, however, that the curves fitted to, say, the combination F_1F_3 shall be derivable from the curves fitted to the frequency combinations F_1F_2 and F_2F_3. In other words, if (i_1,i_2) is a point on the curve fitted to the data from the F_1F_2 combination, and if (i_2,i_3) is a point on the curve fitted to the data from the F_2F_3 combination, then (i_1,i_3) must be a point on the curve fitted to the F_1F_3 data. It will appear that the (tested) assumption of transitivity of the data thus serves the very useful purpose of reducing the uncertainty inherent in fitting curves to the data. Of course, the actual case is more complicated than the example just outlined, since the actual data stem from a greater number of frequency combinations. The basic principle, however, remains the same.

The attempt at a conjoint fitting of curves to all the sets of data was guided by the observation that the data points of a number of these sets seemed to break up into three segments. Within each of these segments the points cluster around a straight line, but the slopes of these linear segments differ. It was further observed that the points of transition between the three linear segments for different matching functions belong, respectively, to the same equal-loudness contours. For these reasons, it was attempted to locate within each set of data four points, each belonging to one of four equal-loudness contours. Two of these points represented the points of transition between linear segments (in those sets of data where such a transition was actually noticeable). The third point represented the maximum loudness level available, and the fourth point represented the "match" obtained by plotting the threshold intensities against each other.

By trial and error, these four points were moved around and finally tentatively located on each plot of the data from the various frequency combinations so that the following conditions were fulfilled: 1) The points did not violate the transitivity requirement, (meaning, specifically, that in whatever combination a given frequency occurred, its intensity coordinate for any one of the four characteristic points would be the same in all diagrams in which this frequency occurred); 2) The points were located as closely as possible to the middle of the scatter range of the points in the corresponding region of the plot; 3) The points were located in such a way that a straight line connecting neighboring points would pass as closely as possible through the distribution of data points in the corresponding region.

This conjoint curve-fitting process is illustrated in Figs. 4-8. The points in squares indicate a transitive fit based on these five groups of data alone. It will be noted that the intensity coordinate for, e.g., the 125 cps tone corresponding to, say, the next lowest point is 56 db SPL in all of the three diagrams in which this frequency occurs. It will also be noted that of the four sets of data shown in Fig. 7, only two seem to conform to the curve drawn. Such departing sets of data would appear now and then, especially for large frequency separations of the tones to be matched. Possibly in these cases the subject switches to a different task as, for instance, matching "density" rather than "loudness"; or the subject may use a different criterion in establishing his loudness matches (as the interpretation of the data to be offered later might indicate).

Obviously some uncertainty is present in the fitting of points and curves as just illustrated; (at the same time it will be equally obvious that the assumption of transitivity of the data has drastically reduced this uncertainty). When all sets of data collected are taken into consideration, the points and lines fitted to the data of Figs. 4-8 have to be slightly altered; this results in a somewhat poorer fit to the data presented. The points in circles, connected by broken lines, represent these modified locations, including some slight relocations to be described in the following.

Construction of Equal-Loudness Contours

On the basis of a transitive fit of four characteristic points to all sets of data, four equal-loudness contours were plotted, each corresponding to one characteristic point. The appearance of these contours in turn suggested minor changes in the location of the characteristic points.

When plotted in sensation level units (db re threshold), it was noted that the equal-loudness contours approximated horizontal straight lines between approximately 80 cps and 1250 cps within a few dbs. Consequently, it was attempted to relocate the characteristic points in such a way that the equal-loudness contours became completely flat within this frequency range. Since the original deviations from this condition were quite small compared to the uncertainty range of the curve-fitting, it proved possible to carry out this relocation without any serious violations of the data.

After this relocation of the characteristic points was completed, it was noted that the three supra-liminal equal-loudness contours increased with a slope close to 20 db/decade above approximately 1250 cps. Again it was attempted to slightly relocate the characteristic points to follow this condition completely; and, again, the uncertainty range permitted this relocation.

Similarly, it was noted that slopes of 20 db/decade and 40 db/decade, respectively, closely fitted the various equal-loudness contours below approximately 80 cps. Again, a relocation of the characteristic points so as to completely follow these slopes proved possible.

The order of magnitude of relocation of points is indicated in Figs. 4-8. However, the difference in the location of the characteristic points in these diagrams also includes the effect of taking the remaining sets of data into consideration.

The resulting equal-loudness contours corresponding to the characteristic points, plotted in sensation level units, are shown in Fig. 9. Since it is assumed that the data distribute themselves linearly between neighboring characteristic points, any other equal-loudness contour can be derived from Fig. 9 by linear interpolation. It should be noted that in many cases some sections of intermediate equal-loudness contours will not have any of the slopes 0 db/decade, 20 db/decade, or 40 db/decade.

Preliminary results from a second subject indicate that 1) the data satisfy the symmetry and transitivity requirements, 2) each set of data can be described as three linear segments, sometimes with different slopes; 3) equal-loudness contours corresponding to the characteristic points consist of linear segments with slopes of either 0 or 20 db/decade when plotted in sensation level units. For this subject, the location of these slopes, and the frequencies at which the slopes change, are quite different compared to Fig. 9.

Discussion

Validity of Equal-Loudness Contours

It must be pointed out that as far as those segments of the equal-loudness contours are concerned which have slopes different from zero, the uncertainty of the data had permitted the fitting of lines with different slopes, or even the fitting of curves. The reason for attempting to fit equal-loudness contours with the particular slopes actually employed will be discussed below.

At the same time, however, it should be stressed that the uncertainty of the data permits only minor changes in the equal-loudness contours in the frequency region from 80 cps to 1250 cps. Within this region an irregular curve with an average slope very close to zero is the only other alternative.

Possible Interpretations of the Equal-Loudness Contours

The fact that the exact transmission characteristic between the eardrum and the inner ear is not known, allows the possiblity of a number of interpretations of the equal-loudness contours of Fig. 9, (and, incidentally, of many other equal-loudness contours consisting of linear segments with slopes of 0, 20, and 40 db/decade).

One possibility of interpreting the equal-loudness contours of Fig. 9 is the following. Let us assume that the threshold sound pressure levels at the eardrum happen to correspond to equal second derivatives of the vibration at some point of the basilar membrane, for instance, at the point of maximum excursion. Under this assumption, the contours of Fig. 9 describe the equal-loudness conditions when the ordinate is regarded as a logarithmic measure of the second derivative of the vibration at some point at the basilar membrane, and it will appear that the middle sections of these contours express that equal loudness are obtained when these second derivatives have the same value.

In this representation, a line with a slope of 20 db/decade represents vibrations on the basilar membrane with the same value of the first derivative, and a line with a slope of 40 db/decade represents vibrations with the same amplitude. Thus, within any single linear segment of the contours of Fig. 9, either amplitude, velocity (which equals the first derivative of the amplitude), or acceleration or pressure (which equals the second derivative of the amplitude) of the corresponding vibrations on some point of the basilar membrane will be constant.

On the basis of these relations we might say that the subject, in a functional sense, for certain frequency-intensity combinations uses the amplitude, for others the velocity, and for the remaining uses the pressure of the basilar vibration as his basis for equating loudnesses. Since no discontinuities seem to occur in the equal-loudness contours, it further follows that at the points of transition from one cue to another, both of these cues furnish the subject with the same loudness-judgment of the tone. It does not appear possible to interpret the particular transition boundaries between the various cues as revealed by this subject.

A number of other basic assumptions may be made, for each of which some linear segment of the equal-loudness contours of Fig. 9 is made to represent some constant aspect of the corresponding vibrations on the basilar membrane. If we wish to restrict ourselves to those assumptions for which the three different slopes represent either amplitude, velocity, or pressure, these assumptions are six in number, of which one has already been described. In two of these other assumptions it is assumed that one complete equal-loudness contour represents a constant aspect of the corresponding vibrations. Thus, Figs. 10 and 11 represent the cases in which velocity and amplitude, respectively, of the corresponding vibrations are assumed constant within one or more equal-loudness contours. Finally, Figs. 12-14 represent more complicated cases.

Each of the six assumptions outlined above implies a different transmission loss as a function of frequency between the eardrum and some point at the basilar membrane. These six transmission characteristics are dipicted in Fig. 15. In case this transmission characteristic had been known in advance from independent studies, a basis for selecting among our six possible assumptions had been present. Since no exact knowledge to this effect is available, however, we have no basis for choosing one assumption rather than another.

Relation to Previous Studies

At first sight, the difference between the equal-loudness contours of Fig. 9 and those, e.g., of Fletcher & Munson (1933) is quite perplexing. In an attempt at resolving this discrepancy, we will examine differences in experimental procedure as well as differences in data analysis between a classical approach and our own.

Our experimental procedure mainly departs from the classical one in that 1) no fixed standard tone (neither with respect to intensity within each run of determinations, nor with respect to frequency from one run to the next) is employed; 2) within each run of determinations, the frequencies of the tones to be compared remain constant. Although it is quite conceivable that these procedural differences may have some effect on the data, they hardly suffice to account for the striking difference between the equal-loudness contours obtained.

The preliminary results obtained from another subject, which were briefly mentioned previously, may furnish a clue to the discrepancy. Although it was possible also in this case to construct equal-loudness contours consisting of linear segments with slopes of 0 db/decade or 20 db/decade, the resulting contours still showed a much greater resemblance to the classical contours than do those of Fig. 9. These preliminary results thus suggest that quite large differences may exist between the equal-loudness contours obtained from different subjects. This, in turn, implies the possibility that averaging sets of data obtained from different subjects may result in a family of equal-loudness contours which depart radically from the contours belonging to a "non-median" subject. Since the equal-loudness contours published by Fletcher & Munson (1933) and by others were all obtained by averaging the data from several subjects, we may thus have indicated a reason for the discrepancy between earlier findings and the ones reported here.

At a later stage, the final results of the present study will be related to theories concerning the transfer properties of the middle ear (e.g., Flanagan, Møller), and implications with respect to various theories for loudness (especially that of Zwicker) will be discussed.

Summary

In this study, loudness matches were collected from one subject for a great number of frequency combinations, extending from threshold to nearly uncomfortable loudness levels. It was found that the "equal-loudness relation" is both symmetric and transitive. Exploiting these properties of the matching relation, curves consisting of three connected linear segments were fitted to the data for each frequency combination. For component frequencies between 80 cps and 1250 cps, all of these segments had a slope of 1 in log-log coordinates. For all component frequencies the upper linear segment was found to have a slope of 1 (in the same coordinates).

On the basis of the curves fitted to the matching data, a set of equal-loudness contours was constructed. It was found that the contours

corresponding to the transition points between the linear segments fitted to the matching data could be represented as a series of connected linear segments with slopes of either 0 db/decade, 20 db/decade, or 40 db/decade when plotted in sensation level units. In the frequency range from 80 cps to 1250 cps the equal-loudness contours (in sensation level units) were found to be horizontal straight lines. Various suggestions as to the interpretation of these results have been offered.

Bibliography

Churcher, B. G. & King, A. J. The performance of noise meters in terms of the primary standard. J. Inst. Electrical Engrs., 1937, 81, 57-90.

Fletcher, H. & Munson, W. A. Loudness, its definition, measurement, and calculation. J. Acoust. Soc. Amer., 1933, 5, 82-108.

Kingsbury, B. A. A direct comparison of the loudness of pure tones. Phys. Rev., 1927, 29, 588-600.

Robinson, D. W. & Dadson, R. S. A redetermination of the equal-loudness relations for pure tones. Brit. J. Appl. Phys., 1956, 7, 166-181.

Ross, S. A technique for producing known sound pressures at the eardrum. Progress Report No. IV, November 1963, Behavior Analysis Laboratory, University of Michigan.

Ross, S. Logical foundations of psychological measurement. Copenhagen: Munksgaard. (In press; to appear during 1964).

Zwicker, E. & Feldtkeller, T. Ueber die Lautstärke von gleichförmigen Geräuschen. Acustica, 1955, 5, 303-316.

Figure Captions

Fig. 1. Experimental set-up for loudness matching. For details see text.

Fig. 2. Equal-loudness matches (dots and circles) for 80 cps and 200 cps. The dots represent matches for a comparison tone of 80 cps, and the circles represent matches for a comparison tone of 200 cps. The crosses represent predicted values calculated from two other experiments both of which employ a standard tone of 125 cps. Points and circles in circles represent "matches" obtained by plotting against each other the threshold values collected within the same experimental session.

Fig. 3. Equal-loudness matches (dots and circles) for 50 cps and 320 cps. Dots and circles represent matches for a comparison tone of 320 cps and 50 cps, respectively. Crosses represent predicted matches derived from two other experiments both with a standard frequency of 125 cps. Points and circles in circles represent threshold "matches."

Figs. 4, 5, 6, 7, 8. Loudness matches for various frequency combinations. Points in squares represent a transitive fit to the data illustrated, and points in circles represent a transitive fit which takes all data obtained in consideration. Circles represent threshold "matches."

Figs. 9, 10, 11, 12, 13, 14. Various representations of the equal-loudness contours corresponding to the "characteristic points" (see text) of the obtained matching functions. In Fig. 9, the ordinate is sensation level (db re threshold). In Figs. 9 and 12 the ordinate may be taken to represent the relative acceleration (pressure) of the vibration on the basilar membrane; in this case, a line with a slope of 20 db/decade represents vibrations of equal velocity, and a line with a slope of 40 db/decade represents vibrations with equal amplitude. In Figs. 10

and 13 the ordinate is assumed to represent the relative velocity of vibrations on the basilar membrane, in which case lines with slopes of +20 db/decade and -20 db/decade represent vibrations with equal amplitude and equal pressure, respectively. Finally, in Figs. 11 and 14 the ordinates depict relative amplitude of vibrations on the basilar membrane; here, slopes of -20 db/decade and -40 db/decade represent vibrations of equal velocity and equal pressure, respectively.

Fig. 15. Transmission loss from eardrum to basilar membrane as a function of frequency. The curves marked P1, V1, A1, P2, V2, and A2 correspond to the assumptions depicted in Figs. 9, 10, 11, 12, 13, and 14, respectively.

Table 1

Frequency Distribution of Frequency Combinations Employed for Loudness-Matching

Lower frequency cps	Frequency Separation 1/5	2/5	3/5	4/5	6/5 decade	
20	1	1	0	0	0	
32	2	1	0	0	0	
50	5	7	1	2	1	
80	6	9	1	0	1	
125	5	5	1	4	1	
200	3	2	2	0	1	
320	3	5	2	5	1	
500	3	3	1	0	---	
800	4	5	1	4	---	
1250	3	2	1	---	---	
2000	3	3	---	---	---	
3200	2	---	---	---	---	
5000	---	---	---	---	---	
	40	36	10	15	5	Total 106

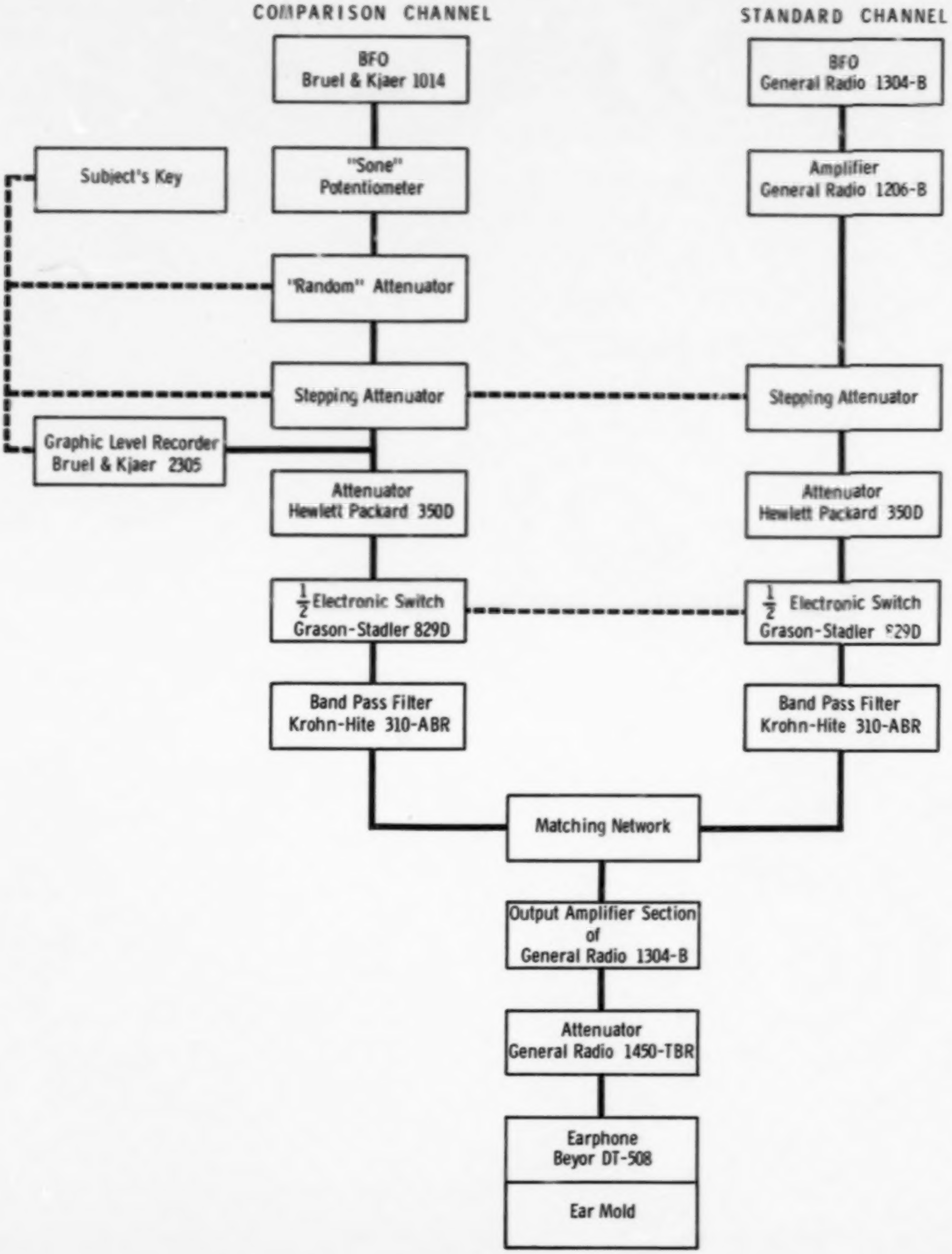
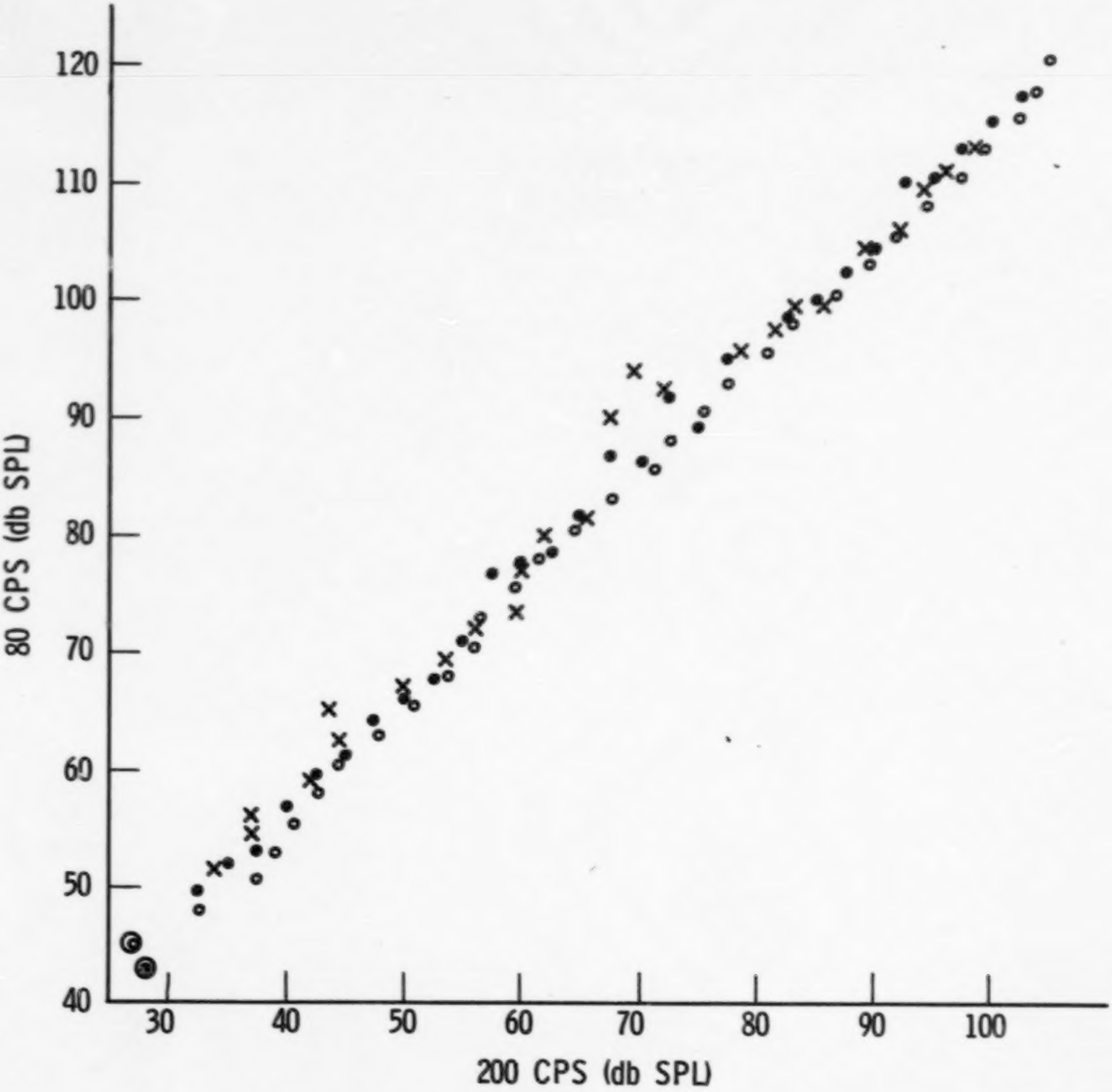
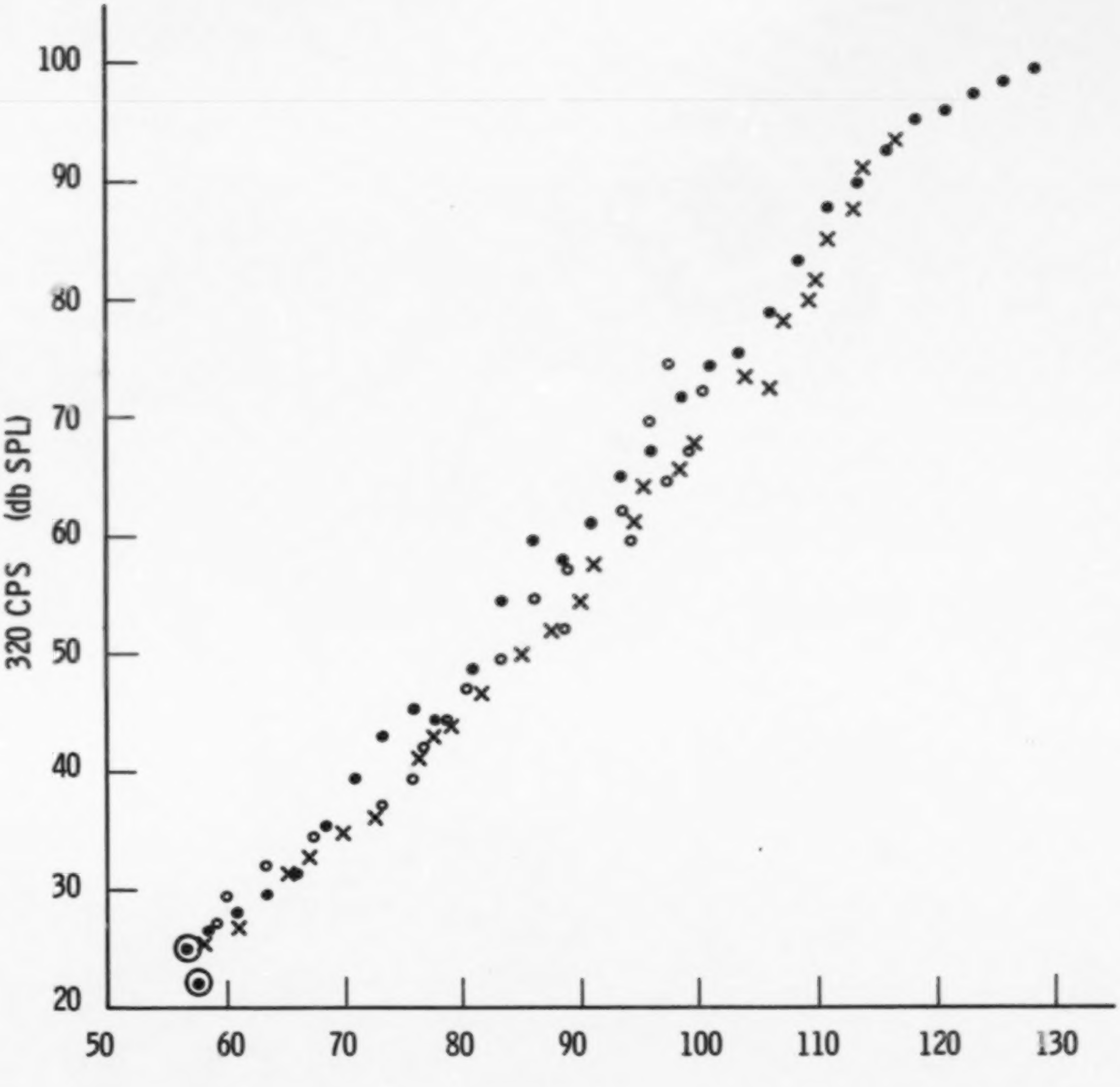

COMPARISON CHANNEL
STANDARD CHANNEL
BFO
Bruel & Kjaer 1014
BFO
General Radio 1304-B
Subject's Key
"Sone"
Potentiometer
Amplifier
General Radio 1206-B
"Random" Attenuator
Stepping Attenuator
Stepping Attenuator
Graphic Level Recorder
Bruel & Kjaer 2305
Attenuator
Hewlett Packard 350D
Attenuator
Hewlett Packard 350D
1/2 Electronic Switch
Grason-Stadler 829D
1/2 Electronic Switch
Grason-Stadler 829D
Band Pass Filter
Krohn-Hite 310-ABR
Band Pass Filter
Krohn-Hite 310-ABR
Matching Network
Output Amplifier Section
of
General Radio 1304-B
Attenuator
General Radio 1450-TBR
Earphone
Beyor DT-508
Ear Mold

Fig. 1



80 CPS (db SPL)
120
110
100
90
80
70
60
50
40
30
40
50
60
70
80
90
100
200 CPS (db SPL)

Fig. 2



320 CPS (db SPL)
100
90
80
70
60
50
40
30
20
50
60
70
80
90
100
110
120
130
50 CPS (db SPL)

Fig. 3

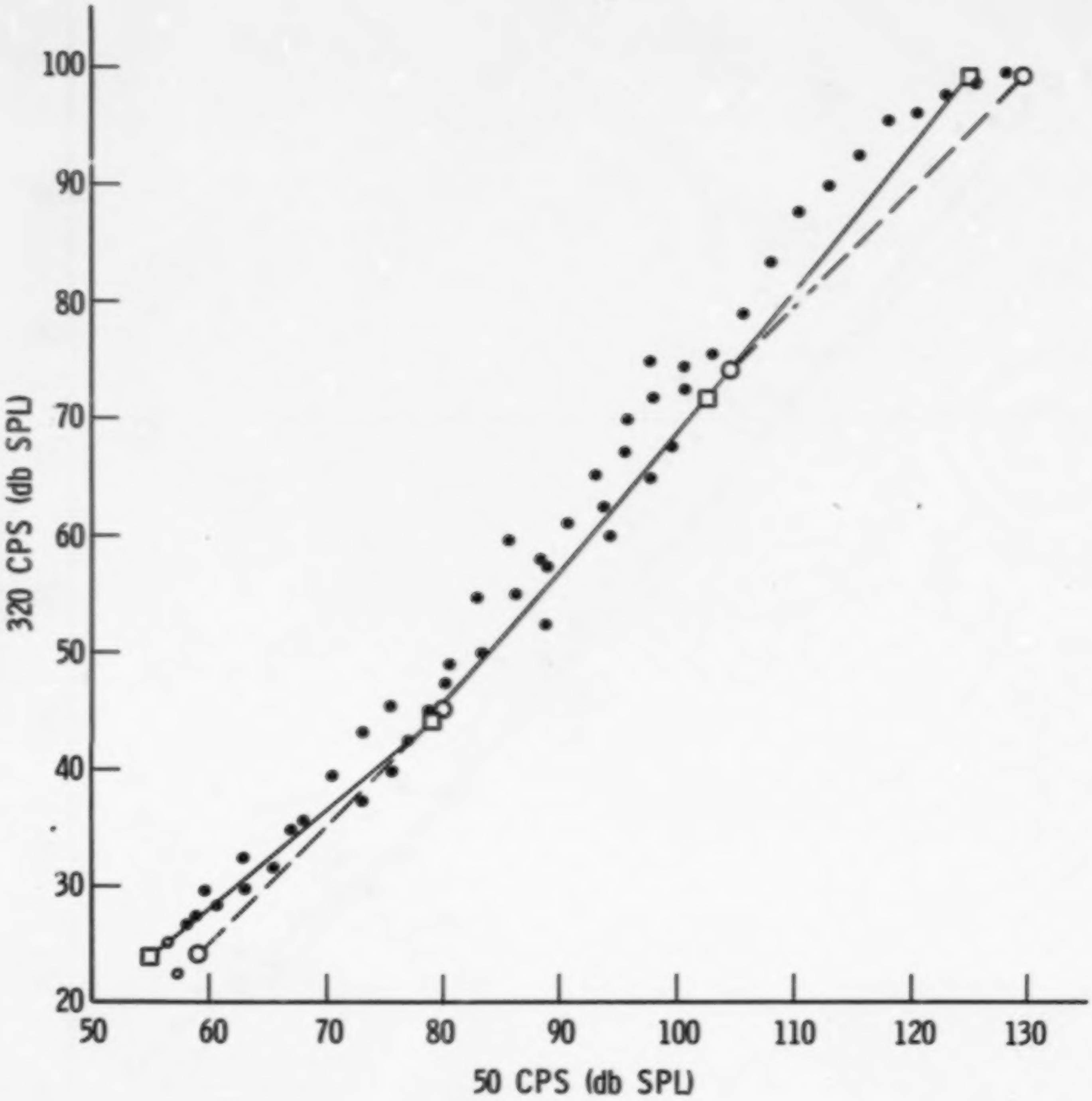
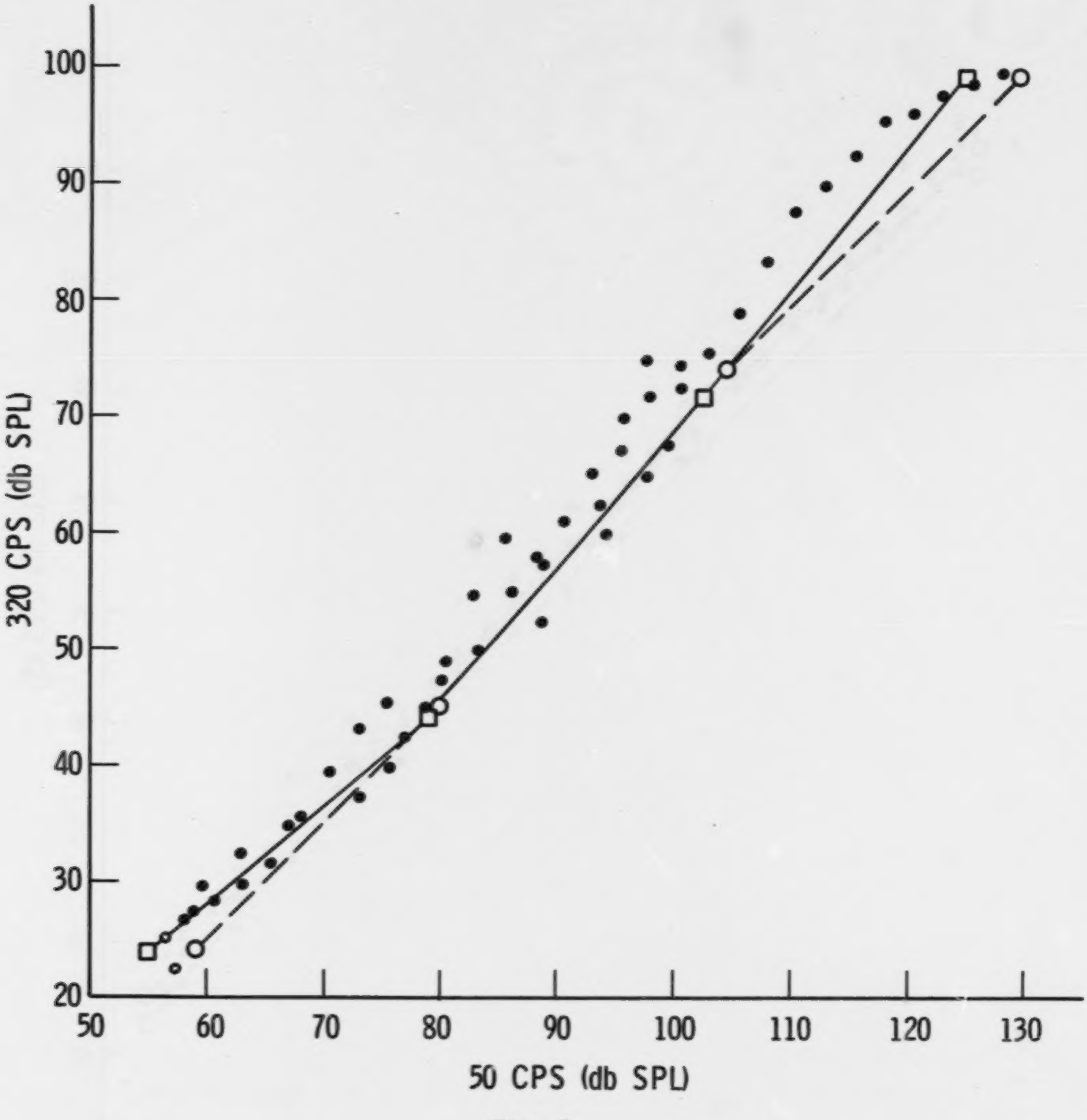
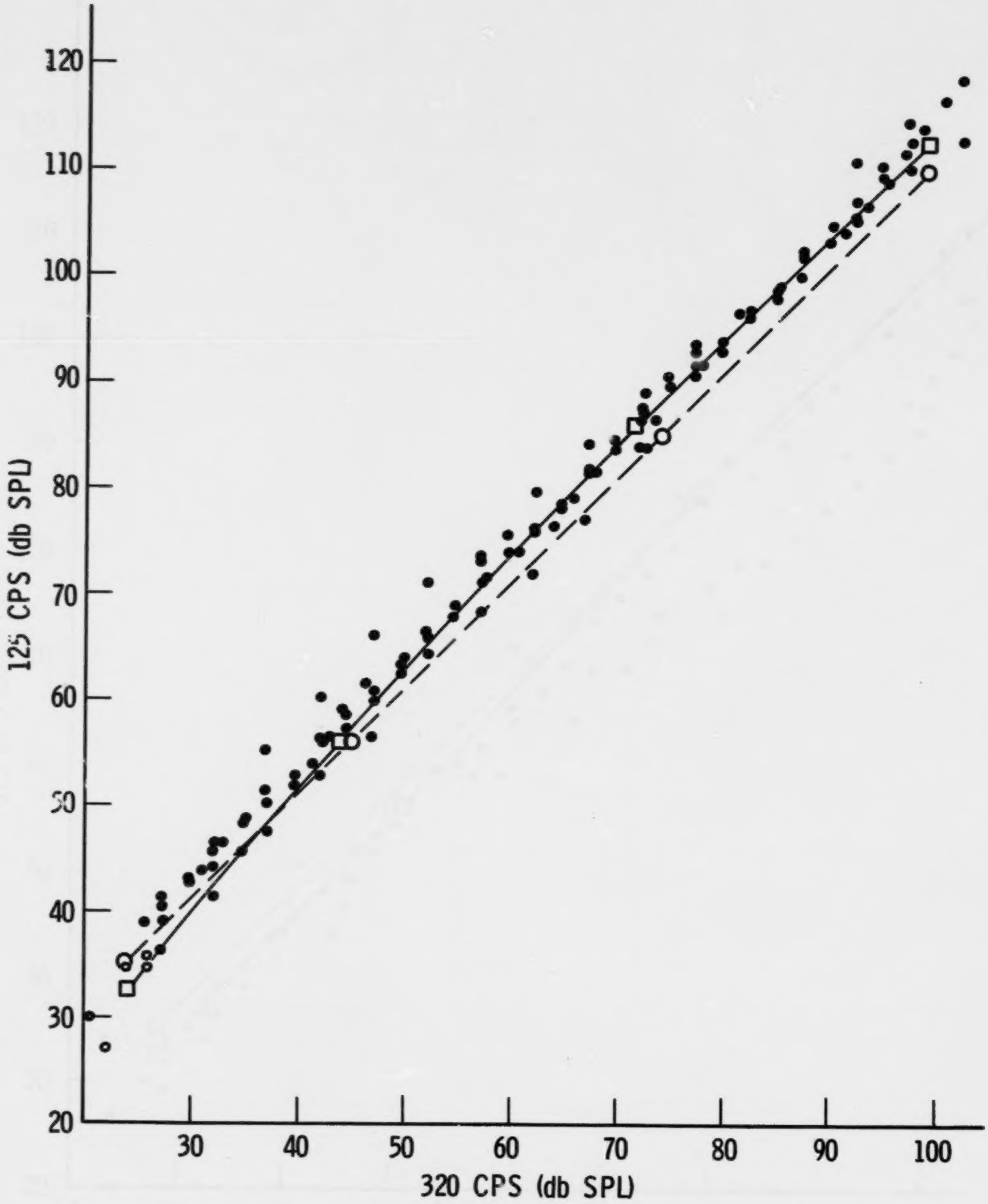

320 CPS (db SPL)
100
90
80
70
60
50
40
30
20
50
60
70
80
90
100
110
120
130
50 CPS (db SPL)

Fig. 4



320 CPS (db SPL)
100
90
80
70
60
50
40
30
20
50
60
70
80
90
100
110
120
130
50 CPS (db SPL)

Fig. 5



125 CPS (db SPL)
120
110
100
90
80
70
60
50
40
30
20
30
40
50
60
70
80
90
100
320 CPS (db SPL)

Fig. 6

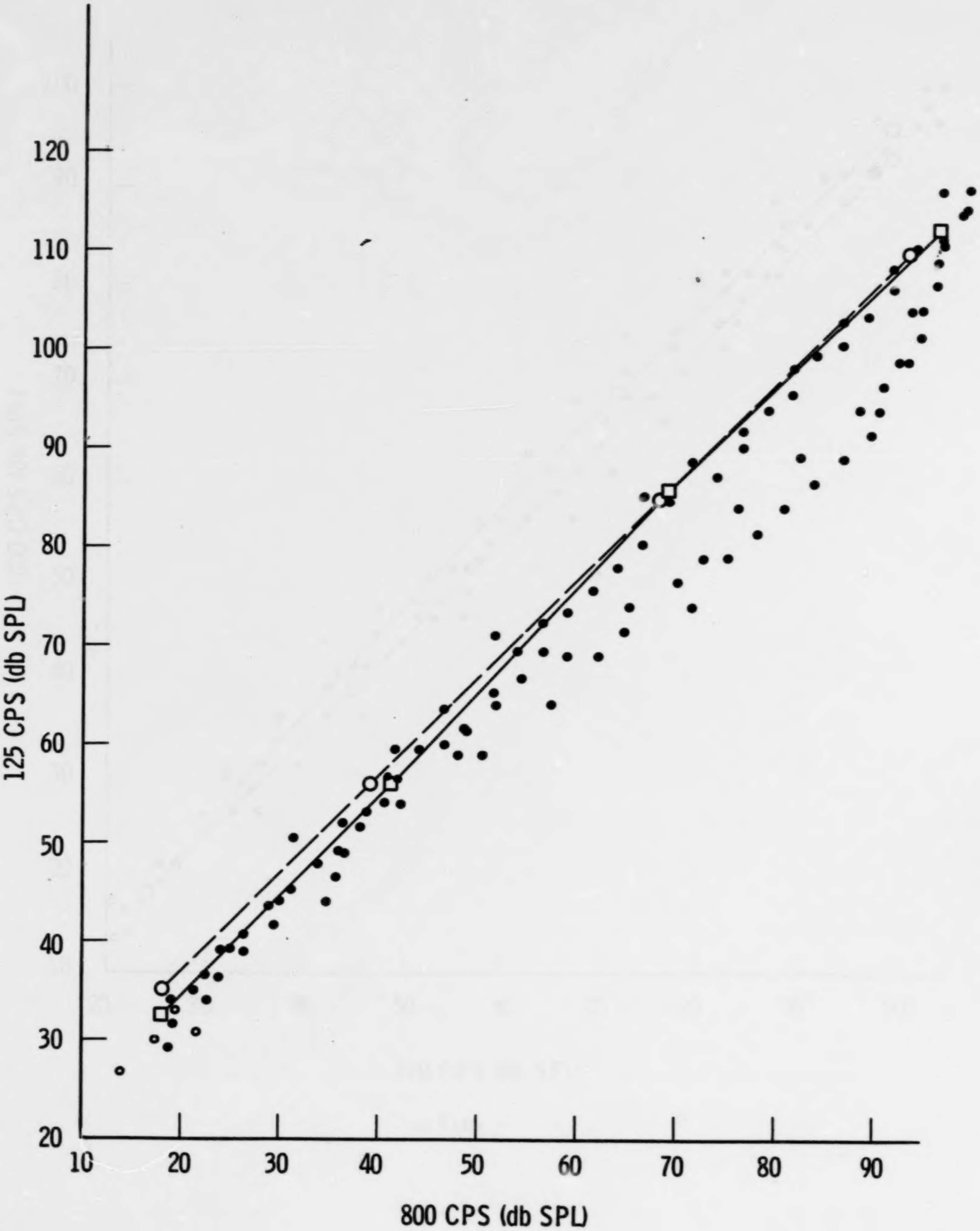

120
110
100
90
80
70
60
50
40
30
20
125 CPS (db SPL)
10
20
30
40
50
60
70
80
90
800 CPS (db SPL)

Fig. 7

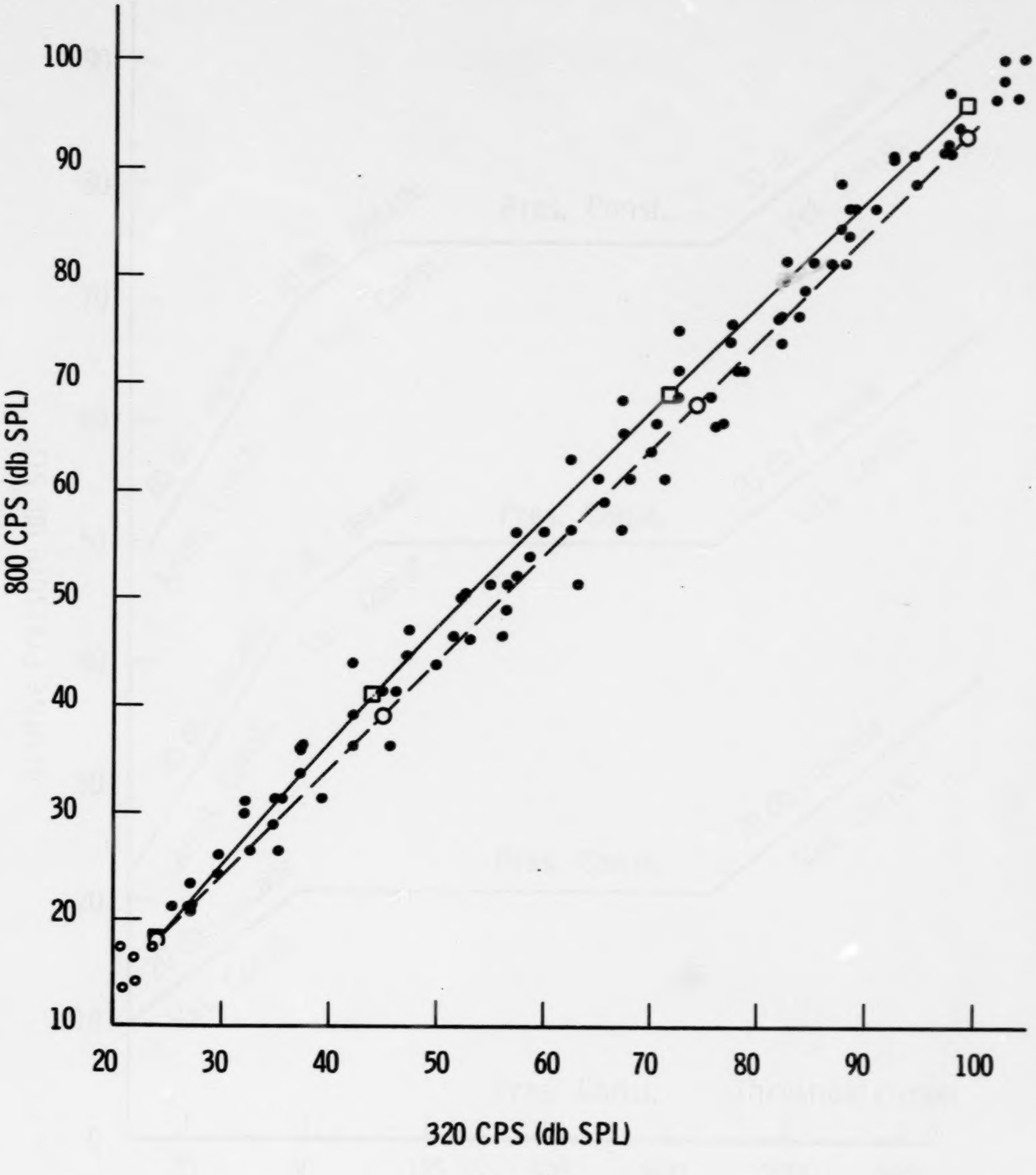
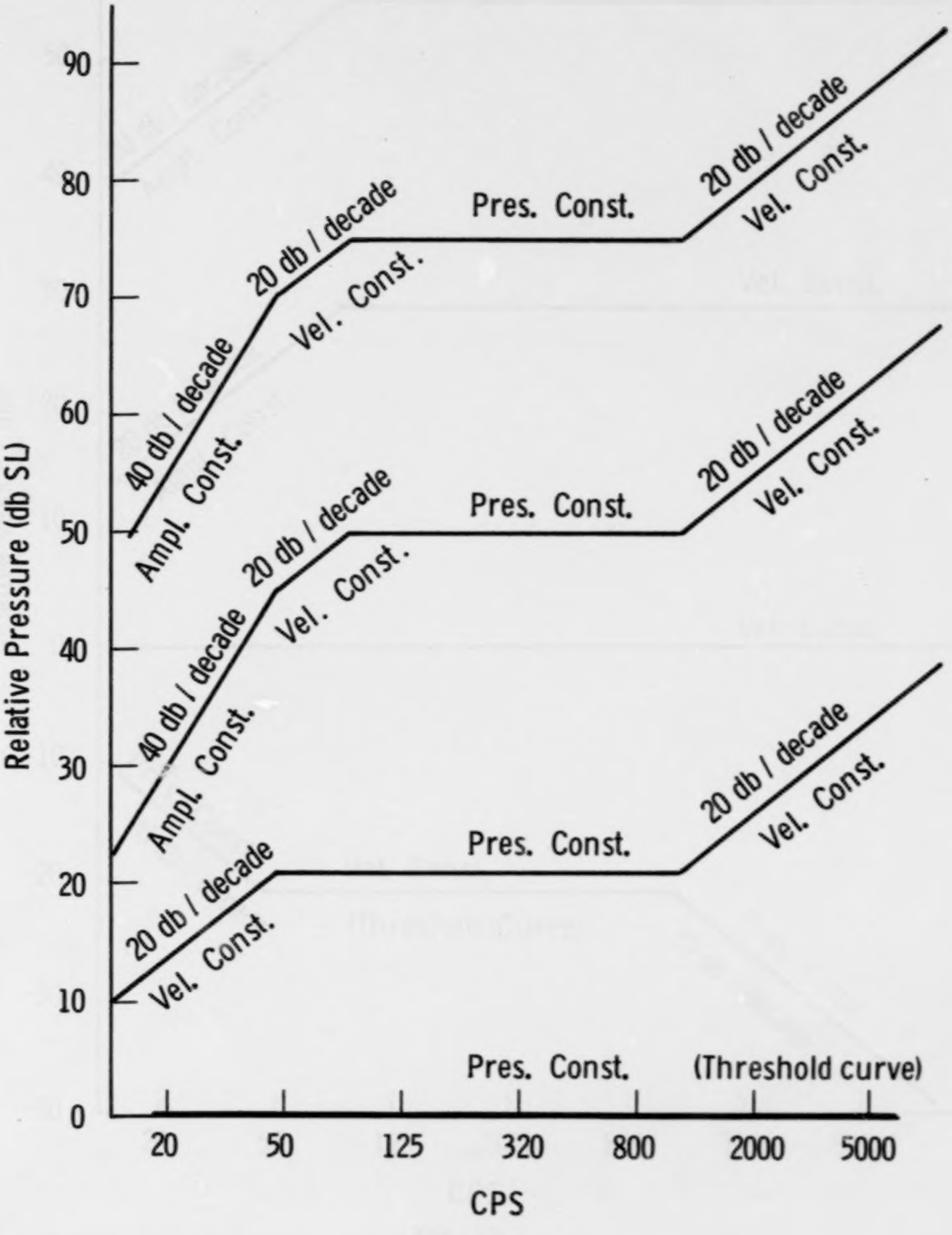

800 CPS (db SPL)
100
90
80
70
60
50
40
30
20
10
20
30
40
50
60
70
80
90
100
320 CPS (db SPL)

Fig. 8



Relative Pressure (db SL)
0
10
20
30
40
50
60
70
80
90
40 db / decade
Ampl. Const.
20 db / decade
Vel. Const.
Pres. Const.
20 db / decade
Vel. Const.
40 db / decade
Ampl. Const.
20 db / decade
Vel. Const.
Pres. Const.
20 db / decade
Vel. Const.
20 db / decade
Vel. Const.
Pres. Const.
20 db / decade
Vel. Const.
Pres. Const.
(Threshold curve)
20
50
125
320
800
2000
5000
CPS

Fig. 9

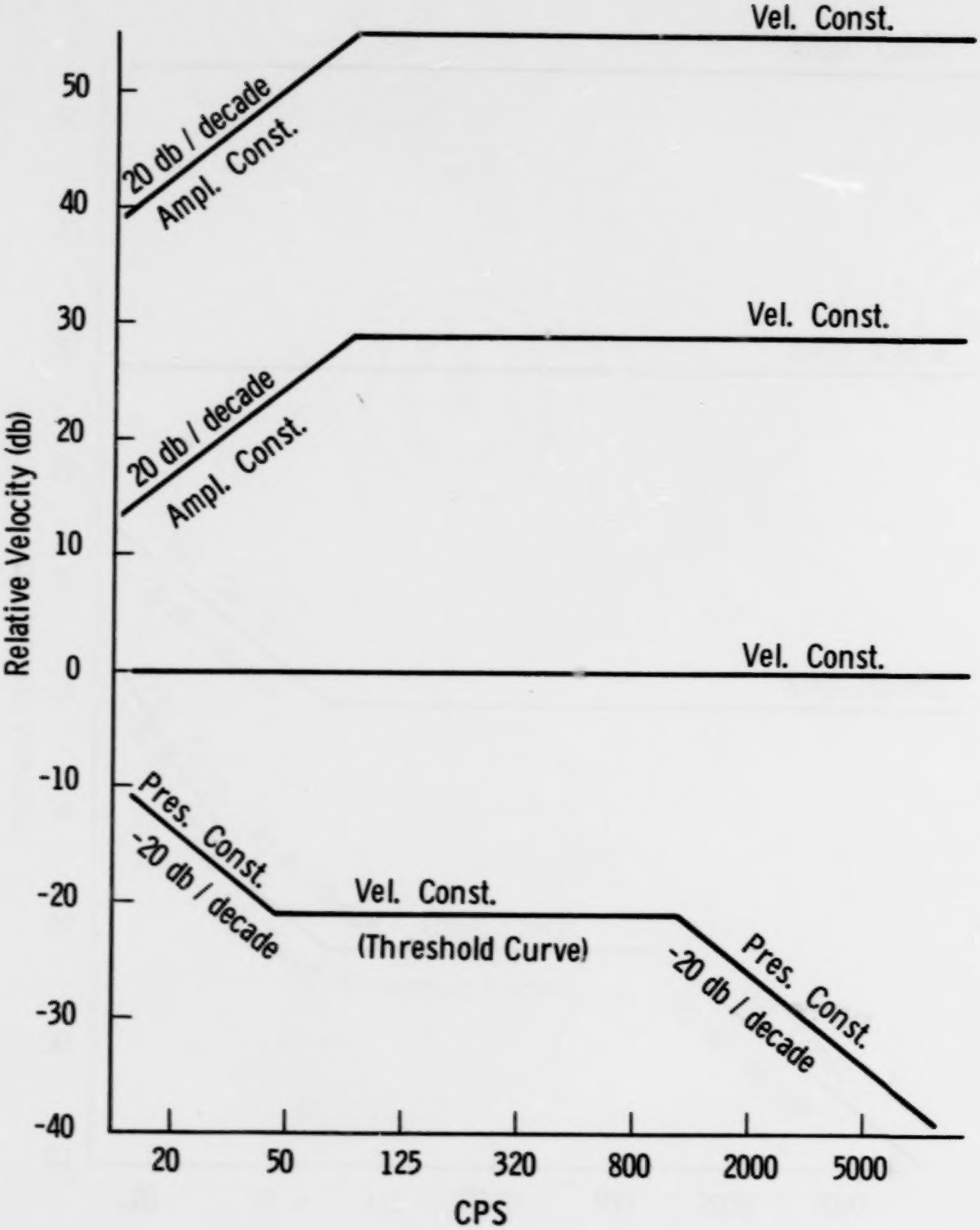

Vel. Const.
20 db / decade
Ampl. Const.
Vel. Const.
20 db / decade
Ampl. Const.
Vel. Const.
Pres. Const.
-20 db / decade
Vel. Const.
(Threshold Curve)
Pres. Const.
-20 db / decade
Relative Velocity (db)
50
40
30
20
10
0
-10
-20
-30
-40
20
50
125
320
800
2000
5000
CPS

Fig. 10

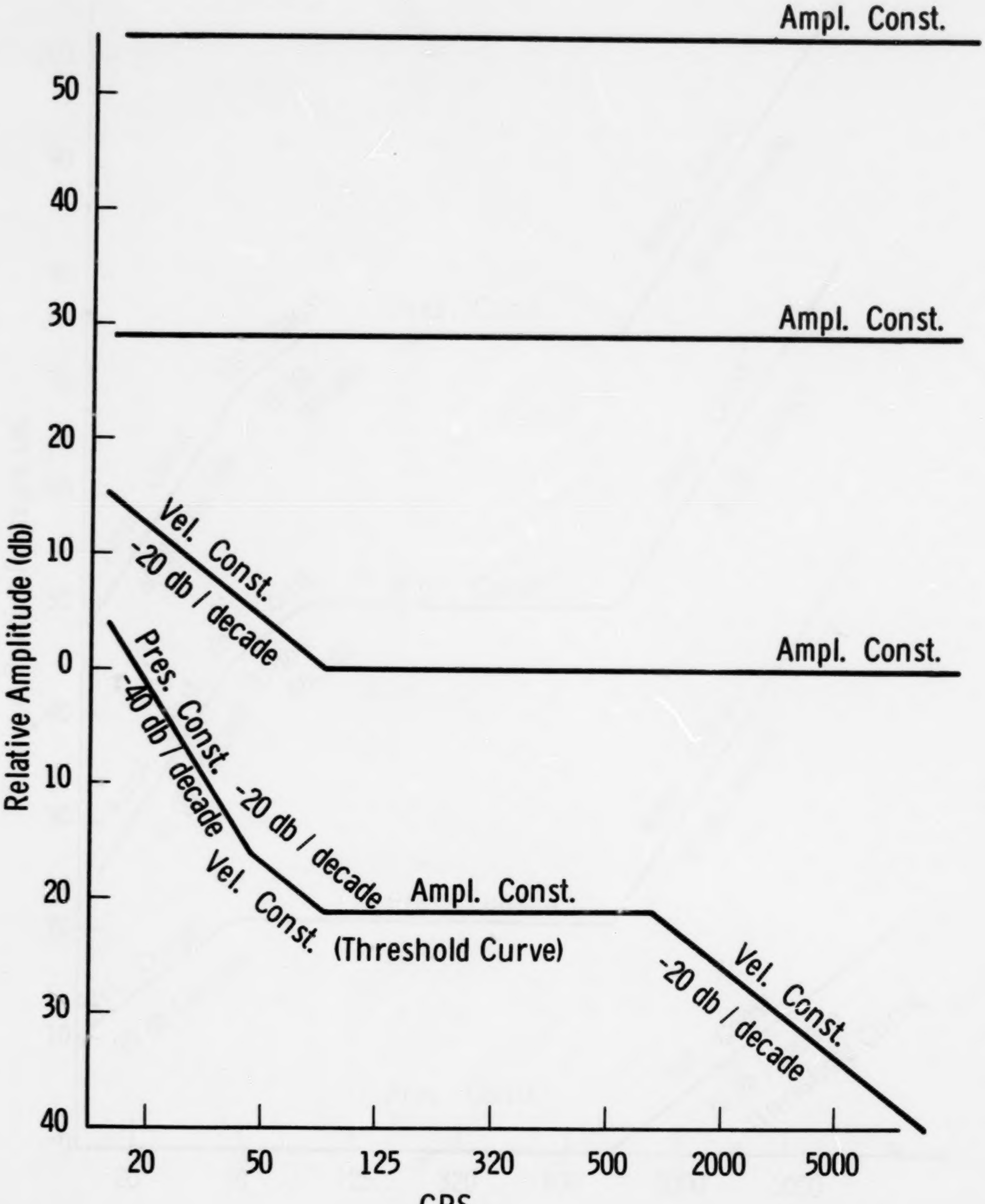

Ampl. Const.
Ampl. Const.
Ampl. Const.
Vel. Const.
-20 db / decade
Pres. Const.
-20 db / decade
-40 db / decade
Vel. Const.
Ampl. Const.
(Threshold Curve)
Vel. Const.
-20 db / decade
Relative Amplitude (db)
50
40
30
20
10
0
10
20
30
40
20
50
125
320
500
2000
5000
CPS

Fig. 11

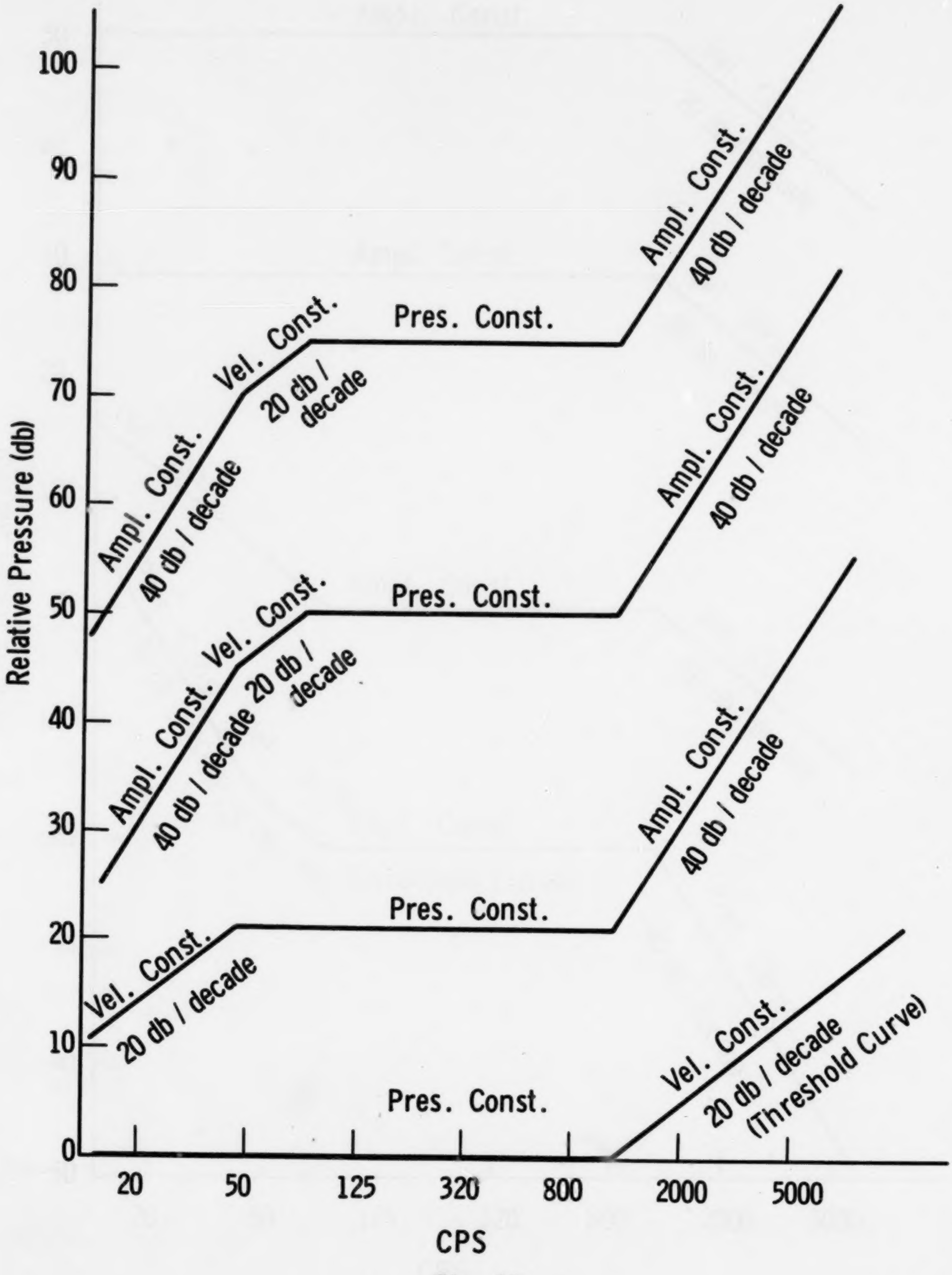

Relative Pressure (db)
100
90
80
70
60
50
40
30
20
10
0
20
50
125
320
800
2000
5000
CPS
Ampl. Const.
40 db / decade
Vel. Const.
20 db / decade
Pres. Const.
Ampl. Const.
40 db / decade
Ampl. Const.
40 db / decade
Vel. Const.
20 db / decade
Pres. Const.
Ampl. Const.
40 db / decade
Vel. Const.
20 db / decade
Pres. Const.
Ampl. Const.
40 db / decade
Pres. Const.
Vel. Const.
20 db / decade
(Threshold Curve)

Fig. 12

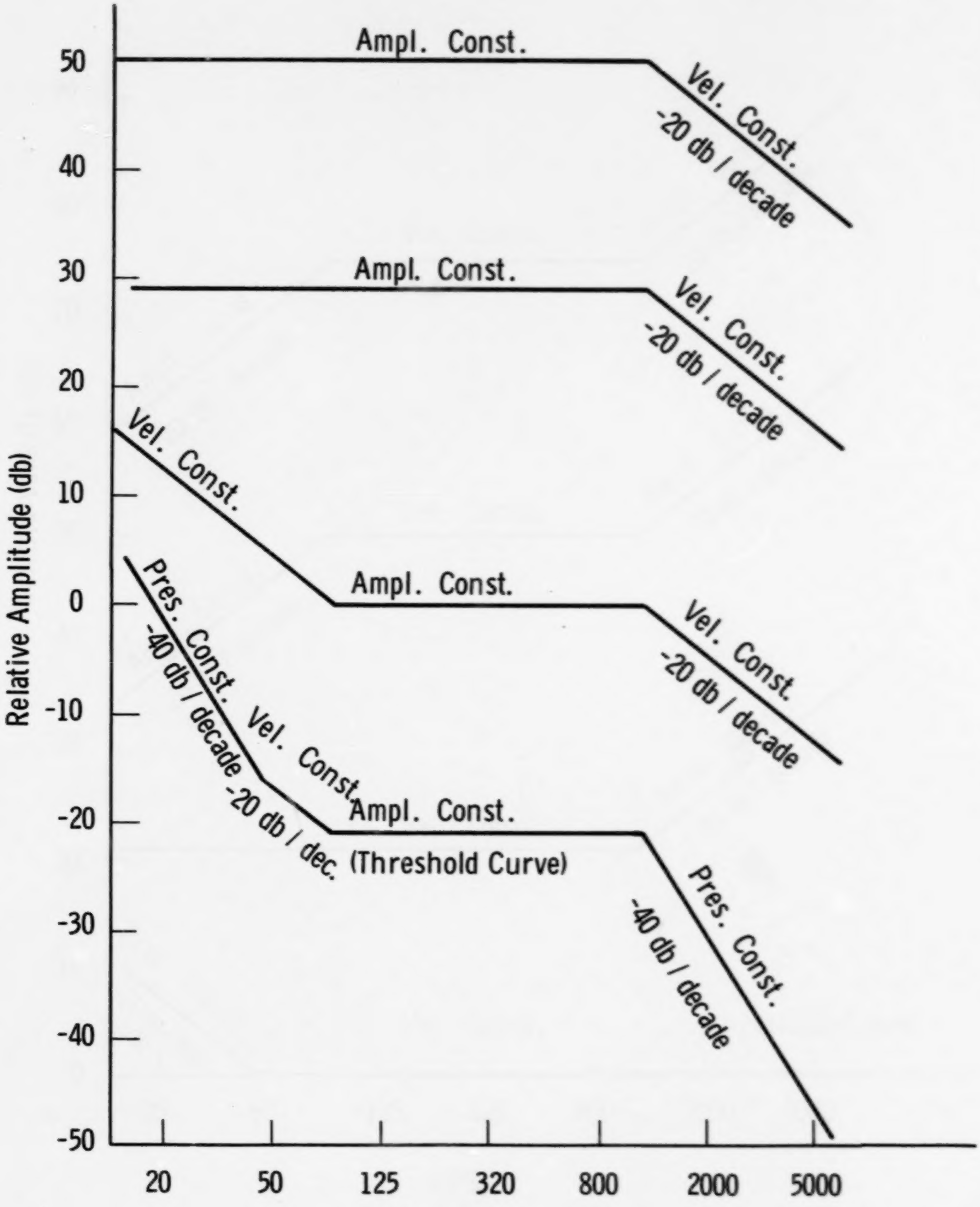

Ampl. Const.
Vel. Const.
-20 db / decade
Ampl. Const.
Vel. Const.
-20 db / decade
Vel. Const.
Ampl. Const.
Vel. Const.
-20 db / decade
Pres. Const.
-40 db / decade
Vel. Const.
-20 db / dec.
Ampl. Const.
(Threshold Curve)
Pres. Const.
-40 db / decade
Relative Amplitude (db)
50
40
30
20
10
0
-10
-20
-30
-40
-50
20
50
125
320
800
2000
5000
CPS

Fig. 13

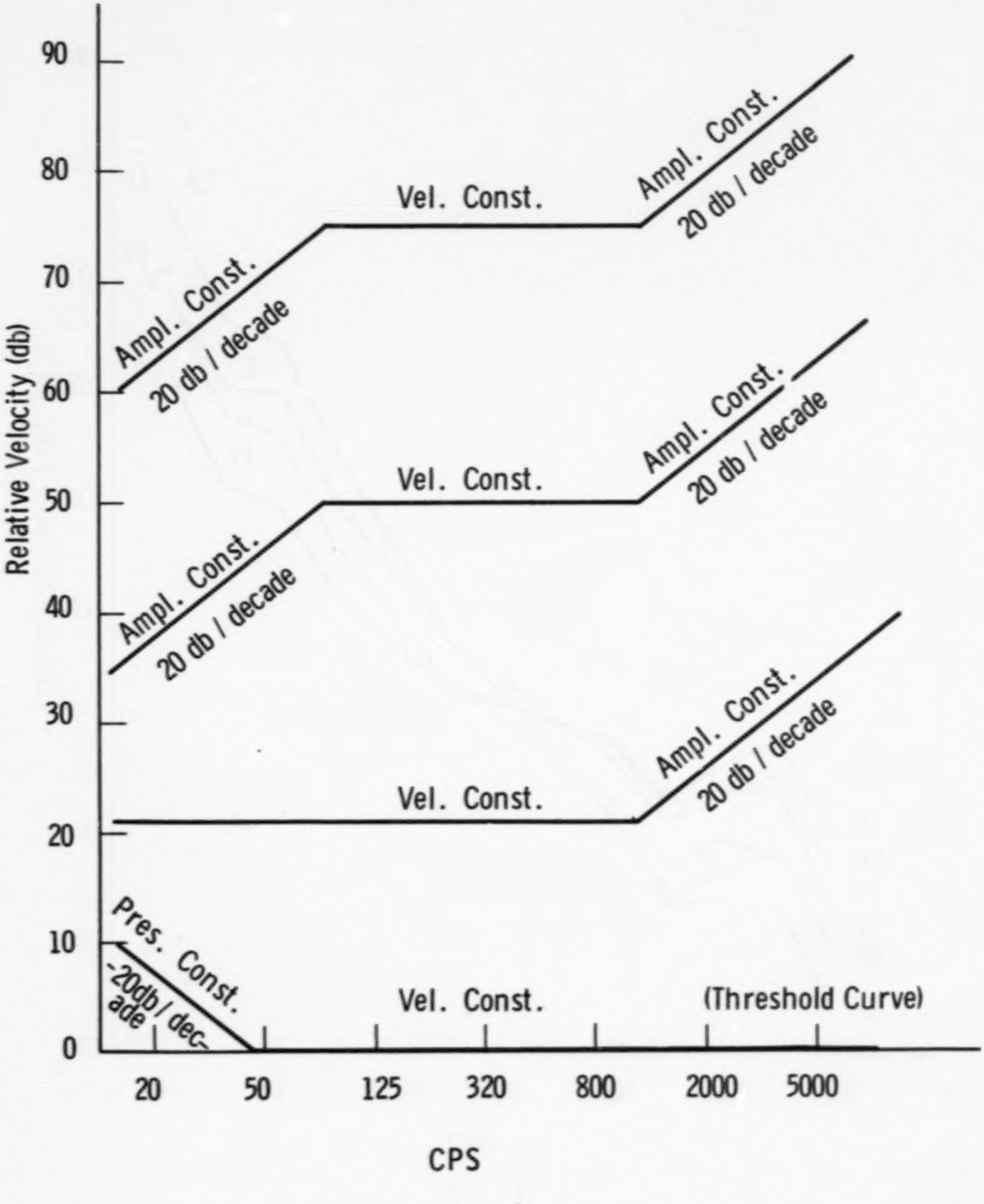

Relative Velocity (db)
90
80
70
60
50
40
30
20
10
0
Vel. Const.
Ampl. Const.
20 db / decade
Ampl. Const.
20 db / decade
Vel. Const.
Ampl. Const.
20 db / decade
Ampl. Const.
20 db / decade
Vel. Const.
Ampl. Const.
20 db / decade
Pres. Const.
-20db / decade
Vel. Const.
(Threshold Curve)
20
50
125
320
800
2000
5000
CPS

Fig. 14

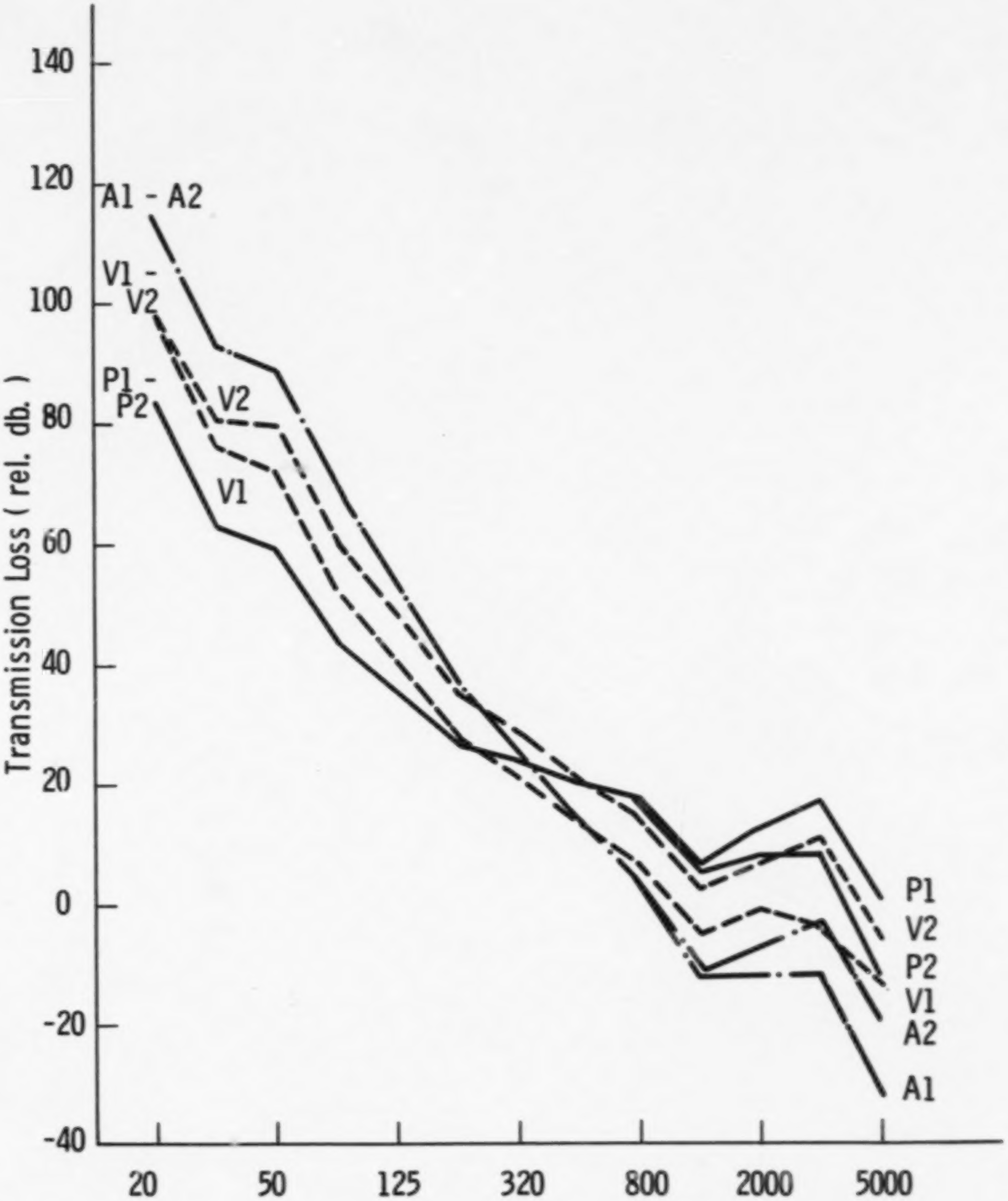

Transmission Loss (rel. db.)
140
120
100
80
60
40
20
0
-20
-40
A1 - A2
V1 -
V2
P1 -
P2
V2
V1
P1
V2
P2
V1
A2
A1
20
50
125
320
800
2000
5000
CPS

Fig. 15

AN ANALYSIS OF THE RELATIONS BETWEEN IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR SPEECH AND NONSPEECH CONTINUA

D. V. Cross and H. L. Lane
The University of Michigan

An Analysis of the Relations between Identification and Discrimination Functions for Speech and Nonspeech Continua

D. V. Cross and H. L. Lane[1]

Behavior Analysis Laboratory

In several studies involving a variety of stimulus continua it has been found that when two stimuli sampled from the same continuum evoke different identification responses they are more readily discriminated than are stimuli (separated by the same number of physical units) that evoke the same identification response.[2] Thus the discrimination functions for these continua are not monotone with the physical stimulus but instead exhibit discontinuities at points along the continuum where a response characteristically evoked in identification of stimuli on one side gives way to a different and sometimes opposite response evoked in identification of stimuli on the other side. For example, an important acoustic cue for the perception of the voiced stops /b/, /d/, and /g/, is the direction and extent of the second-formant transition. When a series of synthetic speech sounds, varied in small steps along this continuum, is presented to listeners who are asked to identify the sounds as /b/, /d/, or /g/, their responses normally tend to divide the acoustic continuum into three sharply-defined phoneme categories (Liberman et al., 1957). The shifts from one response or phoneme label to another are very abrupt and they define points on the continuum called phoneme boundaries. Discriminations within pairs of stimuli, separated by a fixed distance on this continuum (and arranged in ABX triads), are better when the stimuli in the pair fall on opposite sides of a phoneme boundary than when they fall in the same phoneme category. This result has been reported for stimuli from a variety of speech continua (Lane, 1964) including synthesized vowel sounds and also for nonspeech continua (visual as well as auditory).

---

[1] We wish to acknowledge the assistance of Messers. W. Sheppard, G. Lombard and R. Lehman in the design and execution of these experiments.

[2] Bastian, Eimas & Liberman (1961); Cross & Lane (1962); Cross, Lane & Sheppard (1964); Lane & Moore (1962); Liberman, Harris, Eimas, Lisker & Bastian (1961); Liberman, Harris, Hoffman, & Griffith (1957); Liberman, Harris, Kinney, & Lane (1961).

The differences among the outcomes of the various experiments are primarily in the degree to which discrimination functions can be predicted from corresponding identification functions on the extreme assumption that one can discriminate stimuli only to the extent that he can identify them differently. Liberman *et al*. (1961) contend that there is in this regard a qualitative difference between the perception of consonant sounds in speech and the perception of vowel sounds. For consonants, the relation between phonemic labelling and discrimination is apparently very close, whereas the discrimination of vowels is much greater than can be predicted from phonemic labelling functions alone; this difference implies that intra-phonemic differences can cue discriminations among variants of vowel sounds but not of consonants. Liberman interprets this difference as evidence for the theory that "speech sounds are perceived by reference to the articulatory movements that produce them" (Liberman *et al*., 1963, p. 6). This theory maintains that the discrimination of most consonants is discontinuous because they covertly evoke discontinuous (i.e., discretely different) articulations, whose sensory feedback controls the overt identification, thereby mediating the discrimination. Accordingly, the discrimination of vowels is presumed to be continuous because the mediating articulatory movements are continuous.

Actually, the differences between these two types of continua are not as straightforward as the above theory implies. There exists contradictory findings in the case of vowels. On the one hand, Fry *et al*. (1962) found no discrimination peaks at phoneme boundaries, but in their study all stimulus comparisons resulted in perfect or near perfect discrimination. On the other hand, Liberman, Stevens and Ohman (1963) have reported peaks in discrimination functions for Swedish vowels. In support of the hypothesized difference between the perception of consonants and vowels, however, these authors note that the overall level of vowel discrimination was higher than that predicted from the labelling probabilities. The results of an experiment involving a consonant continuum (Liberman *et al*., 1957), however, showed a similar lack of correspondence between obtained and predicted discrimination probabilities. Thus the evidence supporting a motor theory interpretation of the discontinuities in speech perception is hardly conclusive. The fact remains that discontinuities do occur in perception and that there *is* a close relationship between identification and discrimination (for nonspeech as well as speech continua)--although this relationship appears to depend on the operation of other factors that have yet to be thoroughly analyzed.

Predicting discriminations from labelling probabilities. The formula used by Liberman and his collaborators (1957) to predict discrimination probabilities from labelling probabilities rests upon three assumptions the violation of which, in a given experiment, would result in a weakening of the identification-discrimination relationship:

(1) stimuli are discriminated only to the extent that they are labelled differently.

(2) the same responses are used in labelling stimuli in the ABX task as in the identification task.

(3) the probability that response R_j is evoked by stimulis S_i is independent of the responses evoked by other stimuli in the triad.

The first assumption states that perception is categorical, that is, the stimuli are responded to in absolute terms. This means that comparative judgments (A is more like X than B with respect to some property) do not mediate the discrimination. Suppose, for example, that color patches are presented to a subject who is asked to label each as "green," "yellow" or "red." If the same objects are presented in triadic arrangement for the measurement of ABX discriminations, a pair of objects, both of which have been labelled "green," may nevertheless be discriminated perfectly. The restricted response set used in the identification task does not permit assessment of the subtler distinctions that a subject can make among such objects.

The second assumption asserts that the responses involved in the identification task exhaust all responses naturally evoked by the stimuli or, at least, exhaust all responses that are covertly evoked as labels for the stimuli presented in the ABX task. For instance, the color patches in the example above may include some that may be covertly labelled as "blue," "orange," etc. in the ABX task thus mediating discriminations that could not be predicted from the estimated probabilities for the limited set of responses involved in the identification task. Moreover, in an experiment that involves complex, multidimensional stimuli such as speech sounds, the controlling relations that obtain between stimuli and responding may not be the same from one presentation of a stimulus to the next. One aspect or property of a complex stimulus may evoke one response on a given presentation, but on a later occasion the subject may attend a different aspect of the stimulus and, accordingly, respond to it differently. If these additional stimulus properties are correlated with the property controlled by the experimenter, the ABX discriminations will be enhanced; if they are not, then the level of discrimination will be reduced and the assumption that the subject is attending only the relevant stimulus property (and using the responses involved in the identification task) will lead to an overestimate of discriminative accuracy.

The third assumption concerns the effect of context on the probability that a stimulus will evoke a given response. For example, a tone may be judged "loud" when it is the topmost stimulus in one series and "soft" when it is the bottom stimulus in another series. If the probability that a stimulus is labelled R depends on the other stimulus paired with it in the triad, and on the response evoked by the other stimulus, then response probabilities are not independent and the predictive equations employed by Liberman and coworkers are invalid.

Thus a failure to predict discrimination functions from identification functions in the case of vowel continua--and relatively more success in the case of consonants--does not necessarily imply that vowel perception is continuous and consonant perception discontinuous or categorical. The large discrepancy between obtained and predicted discrimination probabilities in the study of vowels by Fry et al., (1962) may be accounted for by a violation of assumptions underlying the prediction. These investigators report that a large context effect was obtained in the labelling of the vowel stimuli. For example, some stimuli were predominantly labelled /ɛ/ when paired with another that was labelled /I/, but they were predominantly labelled /I/ when paired with another stimulus labelled /ɛ/. The question also arises whether a sufficient labelling repertory was provided in the identification task. As a case in point, the same synthetic vowel stimuli employed in the Fry experiment were presented to a subject in our laboratory who was asked to label them as /i/, /I/, /ɛ/, /æ/ or /a/ (Fry et al. permitted their subjects to use only the middle three phoneme labels). Our subject was also instructed to indicate if a vowel did not properly belong to any of these five categories, but rather to some unlabelled category located in one of the four intermediate positions. The outcome was that all five phoneme labels were used as well as the four "intermediate" categories. The subject asserted that he did not use the intermediate categories to indicate indifference or conflict between the two categories that bounded it, but only when it seemed to him that some intermediate phoneme label was in order. The same subject, upon subsequent ABX testing, discriminated all stimulus comparisons perfectly. Thus, assumptions (2) and (3) underlying the predictive formula were violated by the conditions of the experiment by Fry and co-workers. It is also noteworthy, that the tape recordings of the vowel stimuli used in the latter experiment were shown by measurements in our laboratory to contain discriminable intensity changes within triads that were correlated with changes in vowel quality in nearly 50 percent of the cases. Thus, it is very likely, that the obtained enhancement of vowel discrimination above that predicted from vowel labelling is in part attributable to an intensity artifact.

The following experiments were performed to investigate the effects of certain operations on the discriminability of synthesized consonants, modified synthetic vowels and pure tones varying in intensity. We sought to analyze the factors that affect the identification-discrimination relationships that obtain with these stimuli.

Experiment I

In the study of consonant discrimination by Liberman *et al*. (1961), subjects were given extensive practice in identification, labelling the stimuli as /do/ or /to/ speech sounds, before they were given the ABX discrimination task. It was in this prior task that data were obtained from which labelling probabilities were estimated. Since the stimuli employed were synthesized with a device called Pattern Playback (Cooper, 1953) which converts handpainted spectrograms into sound, the synthesized syllables had a quality rather unlike natural speech. One purpose of the present experiment, then, was to measure the discriminability of these synthetic stimuli without the speech set introduced in the identification task. This design would presumably permit the subjects to use labels other than /do/ and /to/ in the identification of these stimuli. Whether these labels would lead to consistent discriminative responding on the ABX task would depend on the stimulus properties attended by the subject and how his identifications correlated with these properties. A second purpose of the experiment was to assess whether feedback as to the correctness of the subject's discriminations would improve the accuracy of his discriminative responding.

Stimuli. The stimuli that were used were those tape-recorded triads of synthetic speech sounds employed by Liberman et al. (1961) and described in detail in their report. In brief, the stimuli were seven three-formant spectrographic patterns that varied only in the relative time of onset of their first and second formants: the onset of the first formant was delayed, in 10 msec. steps, from 0 to 60 msec. High-fidelity equipment was used to play back the tape recordings to subjects (Ss) who served individually in an audiometric room.

Experimental conditions. Five separate experimental conditions were examined according to the following outline:

Group I: Discrimination with nonspeech instructions (N=20).

Group II: Discrimination with speech instructions (N=20).

Group III: Discrimination following phoneme labelling of the triads (N=10).

Group IV: Discrimination with correct responses reinforced followed by discrimination without feedback (N=10).

Group V: Discrimination with correct responses reinforced followed by phoneme labelling, followed by discrimination without feedback (N=10).

In all conditions, except those in which feedback was provided, Ss indicated their responses on an IBM response sheet both in the ABX (discrimination) task and the labelling tasks. In ABX tasks with feedback, the S indicated his choice of stimulus A or B in the ABX triad by pressing one of two buttons on a panel in front of him. If his response was correct, a light flashed and a point was added to a counter also located on the panel. The subject was informed that he would earn a penny for each point earned. Two sets of instructions were used in introducing the ABX tasks in the experiment, depending on whether the stimuli were presented as speech sounds or not. The nonspeech instructions were as follows:

> "This is an experiment in tone discrimination. You will hear a series of tones arranged in sets of three. The third tone will always be identical to either the first tone or the second tone in the set. Your task is to mark space one (press button A) if you think the third tone is most like the first tone, or to mark space two (press button B) if you think the third tone is most like the second tone. Be sure to mark a space for each set of three, even if you must guess."

The speech instructions were as follows:

> "This is an experiment in the discrimination of speech sounds. The sounds you will hear are made by a machine to approximate human speech. They will sound like [do] or like [to]. These sounds will be arranged in groups of three. The third sound will always be identical to either the first sound or the second sound. Your task is to mark space one, if you think the third sound is most like the first sound, or to mark space two, if you think the third sound is most like the second sound. Be sure to mark a space for each group of three, even if you must guess."

In addition, the subjects in group I were asked the following questions in order, at the termination of the session:

1. Did these tones sound like anything you ever heard before?
2. What would you compare them to?
3. Did these tones sound like speech sounds to you?
4. Which speech sounds did they sound like?
5. Do you have any comments concerning the experiment?

Results and Discussion

The responses of the Ss in group I to the questionnaire were surprising. Very few gave any indication that these stimuli sounded like speech sounds. Their responses to these questions are given in Table 1.

---

insert Table 1 about here

Discrimination functions for comparisons of stimuli one, two, and three steps removed on the continuum are shown in Figs. 1 and 2 for all ABX sessions. These functions represent the pooled results of all subjects by experimental condition. (Two Ss in group I stereotyped their responses: one consistently chose the first stimulus in each triad, the other, the second; their results are not included.) ABX discriminations measured after the introduction of a speech set (Fig. 1), whether provided by simply telling S that only /do/ and /to/ speech sounds are involved (group II) or by, in addition, having him label the stimuli in each triad as /do/ or /to/ (group III), are remarkably superior to the ABX discriminations of subjects without a speech set. The latter population consists of group I (Fig. 1), group IV, and group V in its first experimental treatment (Fig. 2). The discriminative accuracy of Ss in these groups, for 1-, 2-, and 3-step stimulus comparisons, is just slightly greater than chance, with no evidence of a peak in discrimination at some point on the continuum. On the other hand, the discrimination functions obtained from groups II, III, and the last treatment of group V--that is, under speech instructions--show definite peaks comparable to those reported by Liberman et al. (1961), although the overall level of discrimination is lower than the levels reported in their study. This discrepancy is probably due to the fact that the Ss in the latter study had much more practice labelling the stimuli.

The procedure of reinforcing correct ABX responses appears to have negligible effect on the Ss' discriminative accuracy (group IV). This result was confirmed in a subsequent study, whose findings are not shown here, in which only those stimulus triads containing 3-step comparisons were presented, again with reinforcement of correct discriminations under nonspeech conditions. Up to 30 replications of the stimulus tape, requiring several one-hour experimental sessions, were given each subject, yet only one S learned to discriminate these stimulus pairs more than 80% of the time, and even this S's level of discrimination dropped back to chance when 1- and 2-step stimulus pairs were re-inserted in the sequence of ABX triads.

The pattern of phoneme labelling by the Ss in groups III and V is shown in Table 2. Only the probabilities of the /do/ response to each of the seven stimuli need be shown, since /to/ response probabilities are complementary.

---

insert Table 2 about here

The labelling results for groups III and V were averaged to obtain estimates of the response probabilities used in predicting ABX discriminations under the assumptions stated earlier. In Fig. 3, the obtained ABX discrimination functions are compared with those predicted for the 1-, 2-, and 3-step comparisons. The correspondence is fairly close but the obtained discriminations are overall higher than those predicted, as is the case for vowels Fry _et al_., (1962) and the /b/, /d/, /g/ continuum (Liberman _et al_., 1957). The result contrasts with that obtained by Liberman and co-workers (1961) who employed the same stimuli as in the present experiment. The prediction of consonant discriminations in their study, however, did not involve the overall labelling probabilities, but rather the context-specific labelling patterns obtained for each triad. In the latter procedure, the predicted proportion of correct discriminations for an ABA triad, for example, would be the number of times that triad evoked the labelling patterns (/d/, /t/, /d/) or (/t/, /d/, /t/) plus 1/2 of times it evoked the patterns (/t/, /t/, /t/), (/d/, /d/, /d/), (/d/, /d/, /t/) or (/t/, /t/, /d/), divided by the total number of presentations of that triad. The predictions obtained with this procedure agree with the formula prediction described earlier only in the absence of a context effect--i.e., when the stimuli in the triad are labelled independently.

It is noteworthy that a simple speech set had a dramatic effect on the discriminability of these so-called speech stimuli. This finding may be related to the fact that the only cue distinguishing the stimuli on this continuum occurs in the first 0 to 60 msec. of the 400 msec. duration of the stimulus. If _S_ is not specifically attending this initial part of the stimulus, but instead searches for differences pertaining to some other property of the stimulus, then his discriminations can be little better than chance level. Responses to the questions asked of the subjects in group I reveal that most _S_s could find little basis for discrimination. Two _S_s reported that duration was the relevant cue, one thought that "pitch" was, and yet another subject described a "chirping" sound that provided the basis for his discrimination.

Since each of the four possible ABX arrangements for each stimulus pair appeared twice in the presentation schedule, the reliability of the stimulus-response relations under each of the experimental conditions can be estimated. This was done in the form of a test-retest reliability measurement by computing the product moment correlation between the number of times a correct ABX response was obtained for a given stimulus pair in the first half of the test schedule and the number correct obtained in the second half. To the extent that S is attending the same stimulus property from one presentation to the next this correlation should be high; if S attends different aspects of the stimulus upon different occasions, however, only chance correlation would be obtained. The correlations computed for the ABX tasks under the various experimental conditions are shown in Table 3. Confirming conclusions reached earlier on the basis of the discrimination functions, it is clear that there is a dramatic increase in the control exerted by the auditory stimuli if S is instructed to respond to the sounds as speech stimuli, i.e., as /do/ and /to/.

---

insert Table 3 about here

Experiment II

The study mentioned earlier investigating the identification-discrimination relation for synthetic vowels (Fry et al., 1962) failed to find discrimination peaks at the phoneme boundaries in the identification functions. It was pointed out that the overall discrimination levels were too high to reflect discrimination peaks. Through the kind cooperation of Dr. A. M. Liberman we obtained a tape recording of the sequence of stimulus presentations used for the vowel study cited and collected labelling and discrimination data from a new group of Ss. Our findings essentially replicated those of the earlier experiment. Even when the tapes were played to Ss without telling them that they were intended to sound like vowel sounds, high levels of discrimination were obtained--in contrast to the findings obtained with the synthetic speech continua in Experiment I. We decided to lower the discriminability of the vowels by degrading their formant structure. This was done by re-recording the entire sequence on a second tape and inserting a broad middle formant intermediate to the two formant frequencies originally comprising each vowel. The extraneous formant was synthesized by applying the output of a pulse generator, which produced a signal with fundamental frequency 130 cps and its harmonics, to a filter set to pass frequencies in the band 1100 - 1600 cps and to attenuate frequencies outside this band at 24 db/octave. The resulting composite stimuli were quite unlike human speech, although each contained its original first and second formants. Since the original stimuli varied slightly in duration as well as in amplitude, the modified stimuli were shortened to a constant 200 msec. by means of an electronic switch. The relative amplitude variations were also reduced somewhat by the insertion of the middle formant. These modified vowel stimuli were presented as in Experiment I to three groups of Ss. The five Ss in each group served for approximately 45 min. each day on 5 consecutive days. The Ss in group I were given the ABX discrimination task with nonspeech instructions (the same instructions given group I in Experiment I) on each of the five days. The Ss in group II were given a nonspeech ABX task on the first day, an ABX task with speech instructions (similar to those given group II in Experiment I) on each of the three following days, and an identification task on the fifth day. In the latter task, they were asked to label each stimulus as an /I/, /ɛ/ or /æ/ vowel. The Ss in group III, like those in the other groups, were given the nonspeech ABX task on the first day, but they were given the identification task on the second day, and then three speech ABX tasks on the remaining three days.

Results and Discussion

Discrimination functions averaged over all days of ABX testing are shown in Figs. 4, 5, and 6 for groups I, II, and III respectively. There is slight evidence for peaks in the discrimination functions between stimuli 4 and 5 and between stimuli 7 and 8. Reference to Fig. 7, which shows the relative frequency with which each vowel label was given to each stimulus for the Ss in groups II and III, indicates--at least for group III whose subjects seemed to be more consistent in their labelling--that the phoneme boundaries occur at the points where discrimination peaks developed (cf. Figs. 5, 6). The inconsistent labelling of these stimuli is evidence of the effectiveness of the degrading operation. Only one S (from group III) yielded identification functions that were even nearly categorical. Figure 8 shows, for this S, the identification functions, the discrimination functions obtained for 1-, 2-, and 3-step comparisons, and the discrimination functions predicted from the labelling probabilities. The marked discrimination peaks obtained in the 1- and 2-step comparisons, and the absence of peaks in the three step comparisons supports our initial hypothesis that enhancement of vowel discrimination at the phoneme boundaries was not obtained in the study by Fry et al.,(1962) because the stimuli selected had such a high baseline of discriminability.

Experiment III

The application of most psychophysical methods (e.g., constant stimuli, paired comparison and stimulus rating methods) to the measurement of discriminations along intensive psychophysical continua yield measures that are continuous and monotone with the physical scale. These procedures depend upon Ss' comparative responses to stimuli relative to their ordering on the underlying continuum. For example, S may be asked to indicate whether one stimulus is greater or less than another with respect to some attribute such as loudness, pitch, brightness, etc. There is little opportunity for S to use categorical responses in these judgments unless, for example, a pair of sound intensities evoke the categorical responses soft vs. loud, thus facilitating the "greater" or "less than" judgment in the constant stimulus task.

In this experiment the discrimination of pure tone intensities was measured by the ABX method in order to assess the relative contributions of categorical and comparative responding to the discrimination of intensity differences.

The stimuli involved were 300 msec. duration, 1000 cps tones differing in intensity from 72 to 90 decibels (SPL) in 3 db steps. Each of the seven stimuli was paired with another differing in intensity by ±3, ±6, or ±9 db, and arranged in triadic sequences forming 1-, 2-, and 3-step comparisons. Thus, for the 1-step comparisons the 72 db tone was paired with the 75 db tone, the 75 with the 78, the 78 with the 81, etc. For the 2-step comparisons, 72 was paired with 78, 75 with 81, etc.; and for the 3-step comparisons, 72 was paired with 81, 75 with 84, etc. From each pair an ABX triad was formed in all possible arrangements (ABA, ABB, BAB, BAA) in which the first two were always different and the third was identical either to the first or second. A total of 120 triads were involved in the presentation schedule (15 stimulus comparisons times 4 ABX arrangements times 2 replications on each arrangement). The triads were pre-recorded on magnetic tape and randomized in blocks of 15. There was 0.5-sec between members of the triad and 5.0-sec intervals between triads.

Subjects and Experimental Conditions. A total of 40 Ss served in the experiment. All were paid volunteers, both male and female undergraduate students. None had any knowledge of the purpose of the experiment nor had served previously in experiments in our laboratory.

Procedure. Ss served individually in sessions lasting approximately one hour. They were randomly assigned to two groups, 20 Ss in each. Each session involved two presentations of the stimulus schedule with a 5-min. break between presentations. In the first presentation, Ss were asked either to (a) rank order the intensities of the tones in each triad by indicating the weakest and strongest tone in the triad or (b) label categorically each tone as "soft" or "loud." In the second presentation, both groups were given the same standard instructions for ABX discriminations. During the experiment, Ss were seated in sound attenuating chambers. The stimuli were played on a tape-recorder (Ampex 350) outside of the chamber and presented to S through calibrated headphones. (TDH-39). The following instructions for the first task were read to the Ss in group I:

> "In this experiment we are investigating how well typical listeners can discriminate the order in which tones of different intensities appear. We will present tones in sets of three through these earphones. Your task is to indicate, for each set of three tones, which tone was the weakest and which tone was the strongest. Indicate your choices on this response sheet by marking in the appropriate slot for each set a '0' for the weakest tone in the set and an 'X' for the strongest tone. If you think, for some sets, that two tones in the set are equally weakest, or equally strongest, you may indicate this by marking both of them with an '0' or an 'X,' whichever seems appropriate. But be sure to make both judgments. That is, there will always be at least one tone that is weakest and one that is strongest--and you must indicate as well as you can which is which. You may find some of these discriminations very difficult to make; don't worry, just do the

best you can. It may help to rely on your first impressions because you will have only 5 seconds between sets to make your choices."

The Ss in group II were read the following instructions:

> "In this experiment we are interested in how typical listeners classify tones as 'loud' or 'soft.' We will present tones in sets of three through these earphones. After hearing each set we would like you to label each tone as loud or soft, whichever you think appropriate. You might label all three tones as soft, all three as loud, or as appearing in some order such as soft-loud-soft, or loud-loud-soft, etc. Indicate your responses on this sheet by marking in the appropriate slot a '0' if you would describe the tone as soft, or an 'X' if you think it is loud. Be sure to respond to each tone in every set, even if you are not absolutely certain how to label it. You will have only 5 seconds between sets to make your responses. In general, your first impressions will be the most accurate. So don't worry, just do the best you can."

After completion of the first tasks the experimenter reentered the chamber and read the following instructions for the ABX discrimination task for all Ss:

> "In this part of the experiment, we will again present sets of three tones. We would like you to indicate whether the third tone is more like the first tone or more like the second tone. If you think the third is more like the first, fill in the space marked (1). If the third is more like the second, fill in the space marked (2). Fill in one and only one space for each set of three tones."

Results

The discrimination task performed with the second presentation of the recorded series of tones yielded nearly identical discrimination functions for both groups of Ss, so their data were pooled before displaying them in Fig. 9. Each datum point there represents the percent correct ABX responses in 160 replications of the stimulus comparison involved, pooled over 40 Ss. Discrimination levels are successively higher for the 1-, 2-, and 3-step comparisons as expected, and there is slight evidence for a discrimination peak between the 81 and 84 db stimuli. The subjects' comparative or categorical responses to the triads presented in the first phase of the experiment were used to predict the probability of a correct ABX discrimination for each stimulus comparison in each of three ways.

(a) Prediction under the assumption that response probabilities are independent. The formula

$$\hat{P}_{\text{corr ABX}} \quad \tfrac{1}{2} + \tfrac{1}{2}\,[P(S|A) - P(S|B)]^2$$

was employed. $\hat{P}_{\text{corr ABX}}$ is the predicted probability of a correct ABX response for the AB stimulus comparison; $P(S|A)$ and $P(S|B)$ are the empirically estimated probabilities of labelling stimuli A and B respectively as "soft." Since only two labelling responses were involved, this formula is equivalent to the one used by Liberman et al. (1957, p. 363). These response probabilities were estimated by calculating the number of times each stimulus was labelled "soft" by the Ss in group I relative to the total number of times the stimulus appeared in the presentation schedule. Because of the way the stimulus comparisons were arranged, the individual stimuli did not appear with equal frequency. The number of labelling responses obtained from each S for the various stimuli were as follows: 36 for the 72 and 90 db tones; 48 for the 75 and 87 db tones; 60 for the 78 and 84 db tones; and 72 for the 81 db tone. Thus, pooling the data from all 40 Ss, the estimates for the labelling probabilities were based on a minimum of 1440 and a maximum of 2880 presentations of each stimulus.

(b) Prediction from categorical responses including context effect (i.e., with non-independence of response probabilities). In this method the predicted probabilities of a correct ABX response were estimated from the labelling patterns obtained for each triad in the identification task. Using only the responses "soft" (S) and "loud" (L), there are eight possible ways a subject may label the stimuli in an ABA triad. For the purpose of predicting how S would respond to this triad in the ABX discrimination task, these eight response patterns can be divided into three classes. One class, consisting of the two patterns (SLS) and (LSL) would lead S to the correct ABX response, "A." A second class, consisting of the patterns (SLL) and (LSS) would lead S to the incorrect ABX response, "B." The remaining four patterns (SSL), (LLS), (SSS) and (LLL) would leave S no alternative but to guess and he would, presumably, choose "A" half the time and "B" the other half of the time. Suppose, for example, that out of 100 presentations of a particular AB stimulus comparison in triad form ABA, the following labelling patterns are obtained with the frequency indicated: (SSS) 10 times, (SLS) 80 times, and (SSL) 10 times. The patterns (SSS) and (SSL) would lead to a correct ABX response (by guessing) half the time or 10 of the 20 presentations. The pattern (SLS) would lead to a correct ABX response every time. Thus, a total of 90 out of 100 presentations would lead to a correct ABX response, for a predicted probability of 0.90 correct discrimination.

(c) Prediction from comparative response patterns.

In this method a procedure similar to (b) was employed using the response patterns obtained from the Ss in group I in ordering the stimuli in each triad as weakest (w) or strongest (s). For example, if the stimuli involved in an

ABA triad were rated either (wsw) or (sws) then both patterns would presumably lead to a correct ABX discrimination, and similarly, the patterns (wss) (ssw) would lead to incorrect ABX responses and the patterns (wws) and (ssw) would lead to indifference between A or B, i.e., a chance probability of being correct. Ss were not permitted to respond with the patterns (www) or (sss) since it was stressed that in each triad there would be a weakest and strongest tone. However, since Ss were strictly required to respond only to the weakest and strongest tone respectively, and to apply these responses to all three tones only when ties were apparent, it was possible to obtain patterns such as (-sw) or (ws-) where the dash indicates an omitted response. It was assumed in these instances that if only two stimuli in the triad were ordered as weakest and strongest then the third tone was perceived as intermediate to the two in loudness, i.e., a simple order prevailed for the triad. There were 12 different possible response patterns in this task: (wws), (wsw), (sww), (ssw), (sws), (wss), (ws-), (sw-), (w-s), (s-w), (-ws), and (-sw). Predicting the probability of a correct response to an ABA triad in the discrimination task involved an enumeration of the various response patterns leading respectively to correct, indifferent and incorrect responses as in (b), with the added complication that for incomplete patterns such as (ws-), representing a simple order for the triad, a correct ABA response would depend on whether the stimulus assigned the middle rank was perceived as being closer to A or to B. Two possible outcomes were evaluated (for the [ws-] patterns) by referring to the number of times a (wsw) pattern was obtained versus the number of times a (wss) pattern was obtained. In the first case a correct ABA response can be obtained and in the latter case an incorrect one. A similar analysis was made for the other response patterns and correct ABX discriminations were then predicted as in (b) with the added operation of "weighting" incomplete response patterns according to the relative occurrence of the corresponding complete patterns.

The predicted ABX discriminations resulting from these computational procedures are compared with those obtained in Fig. 9. The results for 1-, 2-, and 3-step comparisons are shown separately. The differences apparent between predictions based on the independence assumption and the predictions that include the context effect indicate that whether a tone is labelled "loud" or "soft" does depend on the tone it is paired with in the triad. It is also apparent that the predictions based on the assumption that ABX discriminations are mediated by comparative responses to stimuli relative to their ordering on the intensity continuum are more in line with obtained discriminations, than are the predictions based on the assumption that the categorical judgments "soft" and "loud" mediate the discrimination. This result does not preclude the possibility of an interaction between comparative and categorical responses in this task. In fact, the peak in the discrimination function predicted from the comparative response patterns in the 1-step comparison suggests, since it occurs at the point predicted from the labelling probabilities, that comparative responses may also be influenced by categorical response tendencies. That is to say that it is easier to order a stimulus pair with respect to intensity if the stimuli in the pair are labelled differently (i.e., as soft versus loud) than if they are labelled the same way.

Discussion

The results of experiments I, II, and III, considered together, suggest the following conclusions concerning the measurement of stimulus discriminations by the method of ABX triads.

(1) Comparative versus categorical responding. The discrimination of stimulus differences along some continua is greatly improved if S is permitted to 'code' variations in the physical stimulus, absolutely, using categorical response labels. Chistovich, Klaas, and Alekin (1961) found that the discrimination of rapid tone sequences was much greater when S labelled the tones in the sequence in absolute terms than when the same tones were simply ordered as low-middle-high pitch. With the complex stimuli used in experiment I, Ss could find no basis for a comparative response and could discriminate accurately only after a basis for categorically labelling the stimuli was provided. In one sense, responding categorically to stimuli represents a more primitive (albeit in some cases more efficient) level of stimulus control than responding to stimuli comparatively. The operation of categorically labelling elements of a stimulus set results in a partitioning of the set into equivalence classes, where the equivalence relation defined on the stimulus members of each class is reflexive, symmetric and transitive. Stimulus differences along psychophysical continua evoke comparative responses only after a long history of training to respond differentially relative to the order of stimuli within a given equivalence class. Comparative responding represents a higher level of abstraction than does categorical responding. Comparative responding to the order of stimuli within a given equivalence class has been demonstrated in animal research by transfer of training in "transposition" experiments (e.g. Spence, 1937). Abstraction of the properties of "loudness," "pitch," "brightness," etc. from variations in the physical properties corresponding to these attributes leads to an ordering of the stimulus members of the relevant stimulus class and, for human observers, arises from a long history of discrimination training with respect to representative elements of the class. The complex acoustic stimuli that evoke speech responses in the case of most consonants, although they may be

ordered on certain physical continua, do not usually occur in the language community except at extreme or isolated positions on the continuum (Harris et al., 1961). Hence, intermediate points on the synthetic speech continuum are unnatural sounds. The language community does not provide discrimination training leading to comparative responses relative to the order of stimuli on this continuum but only to categorical responses with respect to membership in two or more phoneme categories, or equivalence classes. It is understandable, then, that comparative responses would play no role in ABX discriminations, since S has learned only to identify these sounds categorically. The discrimination of vowel sounds, on the other hand, provides a different picture. Psychophysical attributes of complex tones such as timbre and pitch are correlated with variations in the formant structure of vowel sounds--particularly those produced synthetically (Lane, 1962). Comparative responding relative to these attributes, as well as the categorical response labels evoked by sounds having different spacings between their formants, may both play a role in the discrimination of these sounds.

(2) Context effects. The relativity of the controlling relations that often obtain between stimuli and responding have always been a problem in psychophysical measurement. Several decades of research have shown that the control exerted by a stimulus over the behavior of an organism (within the context of a given experiment) depends on a host of interacting factors: the general level of stimulation; the range, spacing, and distribution of stimulus values making up the presentation schedule; the order in which the stimuli are presented and temporal parameters connected with their presentation; background sources of stimulation; previous experience of the organism and learning, adaptation, and practice effects occurring during the experiment; the payoff structure of the experiment; and so on. These factors affect operations involving comparative

as well as categorical responding. The dependence on context of the identifications and subsequent ABX discriminations obtained in all three experiments reported here is evident. The labelling of vowel sounds may be more dependent upon context than the labelling of consonants. The well-known lack of invariance of the acoustic properties of the vowels (Peterson & Barney, 1952; Stevens & House, 1963; Lindblom, 1963) is matched by a corresponding lack of perceptual invariance (Ladefoged & Broadbent, 1957). Accordingly, Fry _et al_. (1962) reported significant context effects in the identification of synthetic vowels. As demonstrated in experiment III of this study, these context effects must be entered into the formulae that predict discrimination for they surely enter into the discrimination itself. Nevertheless, Fry and his collaborators based their predictions on the assumption that the labelling probabilities were free from context effects--in contradiction to what they themselves proved empirically to be false. The apparent disparity between the discrimination of vowels and that of consonants, and the motivation for introducing a motor theory of speech perception to account for this disparity are reduced, when we recognize that the acoustic properties of the vowels make them prey to comparative as well as categorical identification and discrimination.

References

Bastian, J., Eimas, F. D., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. J. acoust. Soc. Amer., 1961, 33, 842 (abstract).

Chistovich, L. A., Klaas, Yu. A., & Alekin, R. O. The role of imitation in discrimination of sound series. Voprosy Psikhologii, 1961, 5, 173-182.

Cooper, F. S. Some instrumental aids to research on speech. In Report of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching. Wash., D. C.: Inst. Lang. Ling., Georgetown Univer., 1953.

Cross, D. V. & Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. J. exp. anal. Behav., 1962, 5, 487-496.

Cross, D. V., Lane, H. L., & Sheppard, W. C. Identification and discrimination functions for a visual continuum: a test of the motor theory of speech perception. J. exp. Psychol., 1964, in press.

Fry, D. B., Abramson, A. S., Eimas, P. D. & Liberman, A. M. The identification and discrimination of synthetic vowels. Lang. Sp., 1962, 5, 171-189.

Harris, Katherine S., Bastian, J. & Liberman, A. M. Mimicry and the perception of a phonemic contrast induced by silent interval: electromyographic and acoustic measures. J. acoust. Soc. Amer., 1961, 33, 842 (abstract).

Ladefoged, P., & Broadbent, D. E. Information conveyed by vowels. J. acoust. Soc. Amer., 1957, 29, 98-104.

Lane, H. L. Psychophysical parameters of vowel perception. Psychol. Monogr., 1962, 76, 1-25.

Lane, H. L. & Moore, D. J. Reconditioning a consonant discrimination in an aphasic: an experimental case history. J. Speech Hear. Disord., 1962, 27, 232-243.

Lane, H. L. The motor theory of speech perception: a critical review. *Psychol. Rev.*, 1964, in press.

Liberman, A. M., Cooper, F. S., Harris, Katherine S., & MacNeilage, P. F. A motor theory of speech perception. In C. G. M. Fant (Ed.), *Proc. Speech Communication Seminar*. Stockholm: Royal Inst. Technology, 1963.

Liberman, A. M., Harris, Katherine S., Eimas, P., Lisker, L., & Bastian, J. An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Lang. & Speech*, 1961, 4, 175-195.

Liberman, A. M., Harris, Katherine S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, 1957, 54, 358-368.

Liberman, A. M., Harris, Katherine S., Kinney, JoAnn, & Lane, H. L. The discrimination of relative onset time of the components of certain speech and non-speech patterns. *J. exp. Psychol.*, 1961, 61, 379-388.

Liberman, A. M., Stevens, K. N., Ohman, S. Identification and discrimination of synthetic speech sounds. Unpublished report, STL-QPSR-2, Royal Institute of Technology, Stockholm, 1963.

Lindblom, B. Spectrographic study of vowel reduction. *J. acoust. Soc. Amer.*, 1963, 35, 1733-1781.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. *J. acoust. Soc. Amer.*, 1952, 24, 175-184.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430-444.

Stevens, K. N., & House, A. S. Perturbation of vowel articulation by consonantal context: an acoustic study. *J. Speech Hear. Res.*, 1963, 6, 111-128.

Table 1

Responses of the Subjects in Group I to the

Questions about the Synthetic Speech Sounds

Subjects	1. Sound like...	2. Compare to...	3. Sound like Speech sounds?	4. If yes, which ones?	5. Comments?
1	Electronic noise	A French 'non'	A nasal sound in French.	'non' in French	I think it was length of signal that was different.
2	Like a person with false teeth saying no or nope.	Same answer as question 1.	Yes	no and nope or woo and wak	
3	frogs	electronic noises	almost	toll and doll	I could only discriminate a few of the tones from the other tones.
4	drums	A musical instrument, trombone	No		If I didn't hear any difference I marked the first column. I heard no difference at all.
5	no	bleeps on a radar screen	No		All sounded the same sometime. The first two sounded the same sometime.
6	Yes, a radio dial that is turned very fast.	Nothing specific	No		Very difficult; duration and pitch were the basis of my discrimination.

Table 1 (Continued)

Subjects	1. Sound like...	2. Compare to...	3. Sound like Speech sounds?	4. If yes, which ones?	5. Comments?
7	No	TV dial tone for a test pattern	No		**They sounded exactly identical sometimes.**
8	No	steam whistle	No		**challenge**
9	an electronic noise	vowels	No, only vowels	a and o	**Very hard to find a difference at first.**
10	Radar bleeps	Radar bleeps	No		**Most of the sounds sounded alike.**
11	No	Buzz on the telephone busy signal	No		**All hard to tell apart.**
12	No	Fire alert	No		**Most of them sounded the same.**
13	Birds	Stillness of the night with crickets and birds making noises.	No		**I think the purpose was to discriminate the chirping sound. It was very hard.**

Table 1 (Continued)

Subjects	1. Sound like...	2. Compare to...	3. Sound like Speech sounds?	4. If yes, which ones?	5. Comments?
14	Science fiction movie noises	Direct dialing tones over the telephone	No		All three sounded identical.
15	No	Railroad train	No		I thought they were all the same sometimes.
16	organ tones	organ tones	No		Most of them sounded alike.
17	No	fog horn	No		They all sounded the same to me.
18	No	Buzzer under water	No		I was having to guess a lot.
19	No	noise	possibly	hard d	They sound pretty much alike.
20	No	a tuning bar	No		I couldn't hear any difference. They all sounded exactly alike.

Table 2

Proportion of /do/ responses

Delay in First Formant Onset (msec.)

	0	10	20	30	40	50	60
Group III	.969	.965	.888	.563	.193	.108	.053
Group V	.931	.950	.853	.519	.130	.065	.061

Table 3

Reliability of discriminations among synthetic speech stimuli under five experimental conditions. The product-moment correlation between correct discriminative responses to 15 stimulus comparisons, replicated four times each, in the first first half and in the second half of the test schedule.

	Experimental condition	r
Group I:	nonspeech instructions	0.13
Group II:	speech instructions	0.85
Group III:	speech instructions	0.83
Group IV:	nonspeech instructions (and feedback)	0.00
	nonspeech instructions	0.26
Group V:	nonspeech instructions (and feedback)	0.02
	speech instructions	0.87

Figure Captions

Fig. 1. Effects of speech set on the discrimination of three-formant patterns. Before ABX testing with stimulus comparisons one, two and three steps removed on the continuum (continuous, dotted, and broken lines, respectively), Ss in group I were told that the stimuli were "tones," Ss in group II, that they were /do/ or /to/, while Ss in group III were asked to label each tone in all the triads as /do/ or /to/.

Fig. 2. Effects of discrimination training and of speech set in the discrimination of three-formant patterns. The legend is the same as in Fig 1. The recorded series of stimuli was presented once to the Ss in groups IV and V with reinforcement for correct responses and then again without feedback. The Ss in group V also received an interpolated presentation of the series, during which they labelled each sound as /do/ or /to/.

Fig. 3. A comparison of the discrimination functions obtained after labelling three-formant patterns as /do/ or /to/, with these functions predicted from the labelling probabilities. The results for groups III and V have been pooled in both the obtained (filled circles) and predicted (open circles) discrimination functions.

Fig. 4. Effects of speech set on the discrimination of degraded synthetic vowels. A tape-recorded copy of the stimulus series employed by Fry et al. (1962) was altered by inserting a constant additional formant between the first and second formants of each vowel. The five Ss in group I were told on day 1 that the stimuli were tones before testing on days 1 - 5. The legend is the same as in Fig. 1.

Fig. 5. Effects of speech set on the discrimination of degraded synthetic vowels. The five Ss in group II were told on day 2 that the

stimuli were the synthetic vowels /I/, /ɛ/ or /æ/ before testing on days 2 - 4. The legend is the same as in Fig. 1.

Fig. 6. Effects of speech set on the discrimination of degraded synthetic vowels. The five Ss in group III were asked on day 2 to label each vowel in all the triads as /I/, /ɛ/ or /æ/ before testing on days 3 - 5. The legend is the same as in Fig. 1.

Fig. 7. Identification of degraded vowels by the Ss in groups II and III. The phoneme boundaries correspond to peaks in the discrimination functions (cf. Figs. 5, 6).

Fig. 8. Identification and discrimination of degraded vowels by one S in group III who gave relatively categorical labelling. The obtained discrimination functions (filled circles) and those predicted from the labelling probabilities (open circles) are shown for comparisons of stimuli one, two, and three steps removed on the continuum.

Fig. 9. Obtained and predicted discrimination functions for the intensity of pure tones arranged in ABX triads. The discrimination functions were predicted: (a) from the labelling probabilities, assuming the independence of responses; (b) from the pattern of labelling responses in triads (context dependent); (c) from comparative response patterns, i.e., rank ordering of tones within triads.

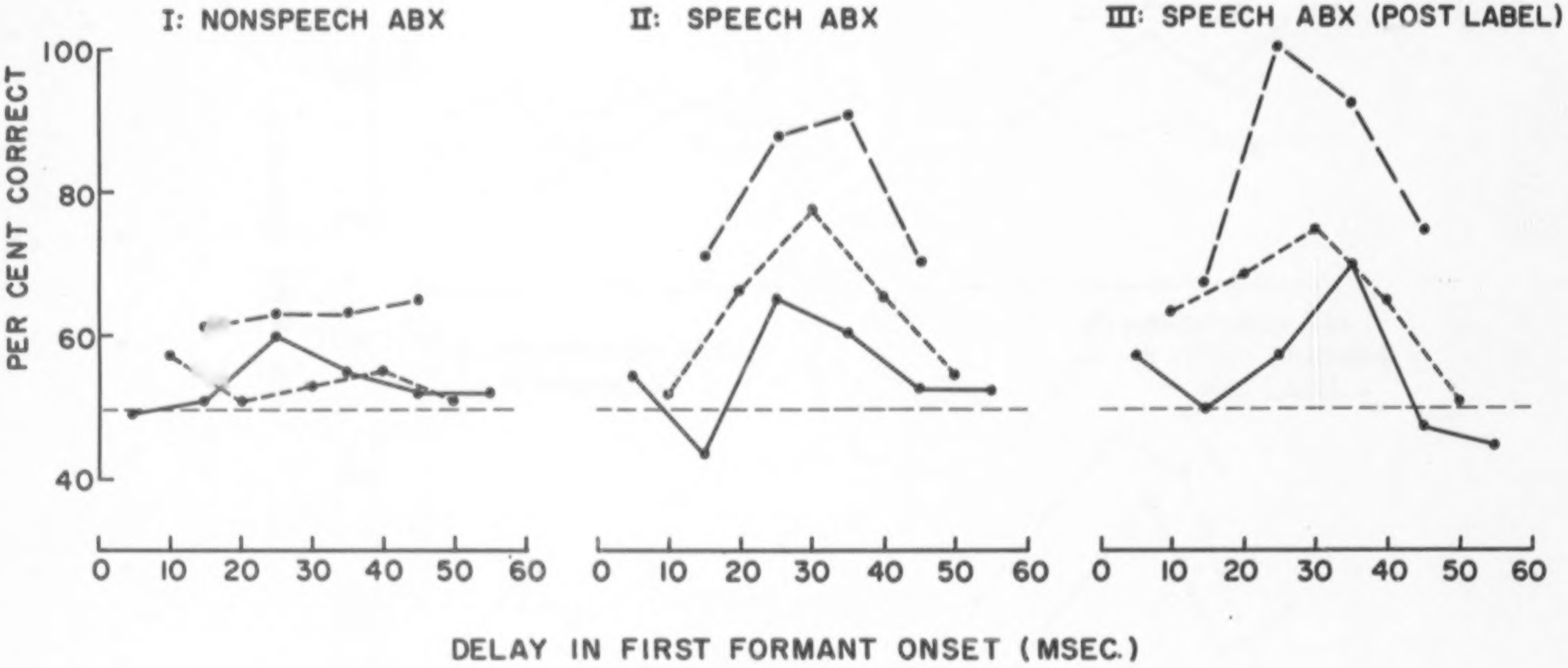

I: NONSPEECH ABX
II: SPEECH ABX
III: SPEECH ABX (POST LABEL)
PER CENT CORRECT
100
80
60
40
0
10
20
30
40
50
60
DELAY IN FIRST FORMANT ONSET (MSEC.)

Fig. 1

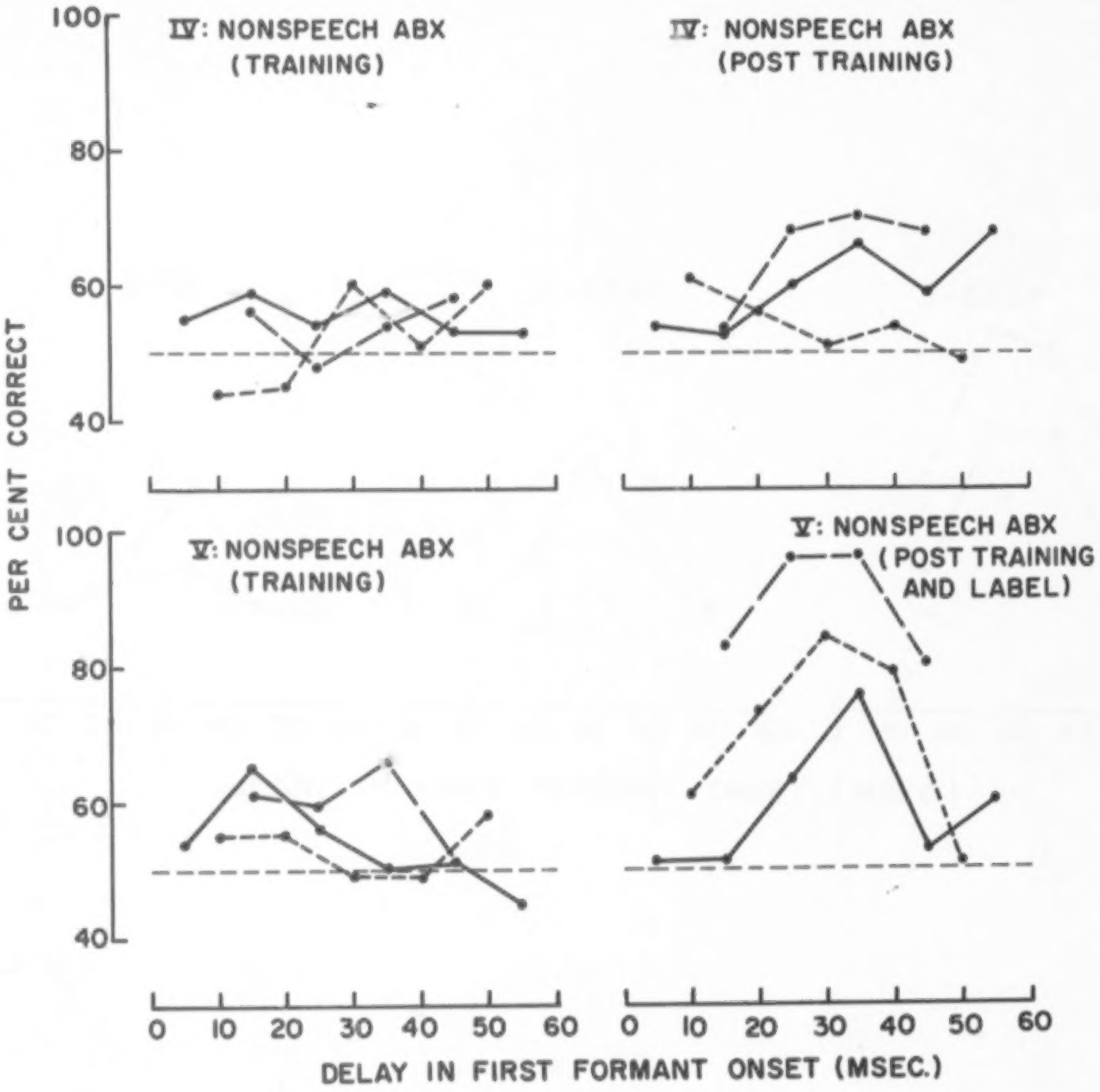

IV: NONSPEECH ABX (TRAINING)
IV: NONSPEECH ABX (POST TRAINING)
V: NONSPEECH ABX (TRAINING)
V: NONSPEECH ABX (POST TRAINING AND LABEL)
PER CENT CORRECT
100
80
60
40
0
10
20
30
40
50
60
DELAY IN FIRST FORMANT ONSET (MSEC.)

Fig. 2

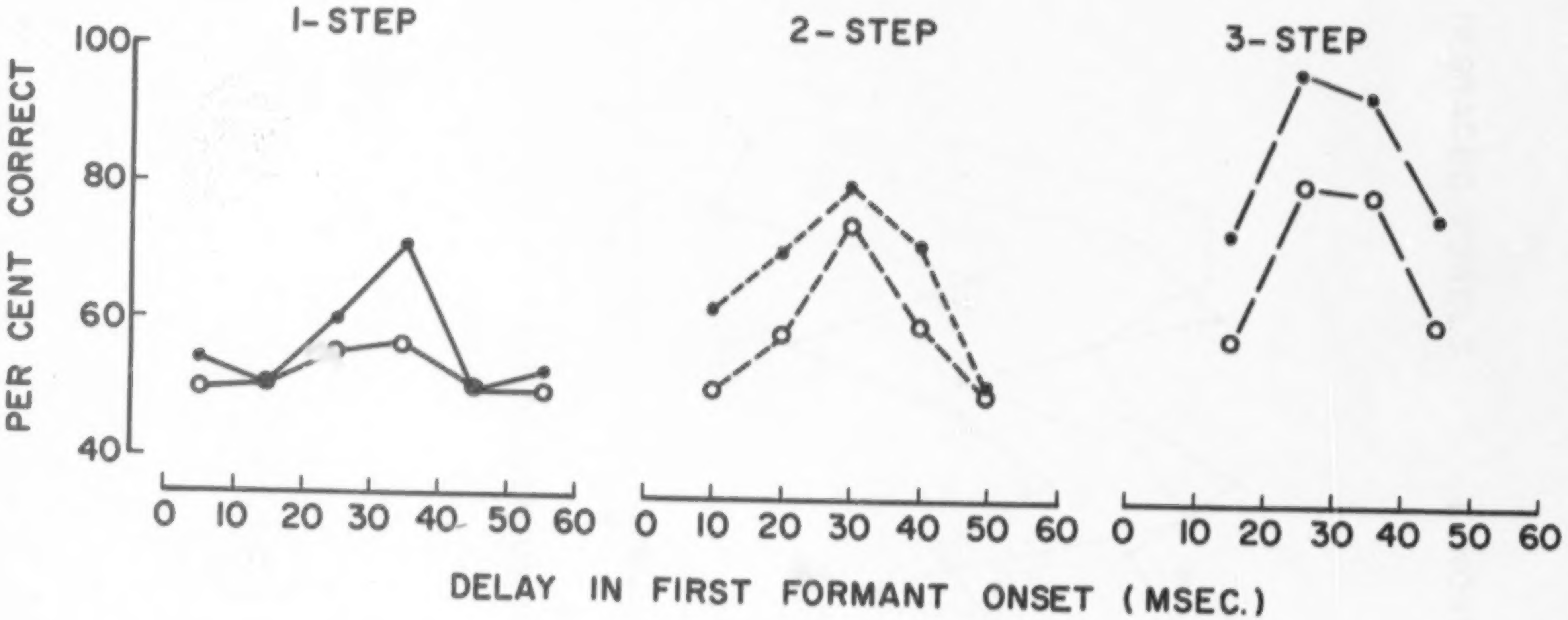

1-STEP
2-STEP
3-STEP
PER CENT CORRECT
100
80
60
40
0
10
20
30
40
50
60
DELAY IN FIRST FORMANT ONSET (MSEC.)

Fig. 3

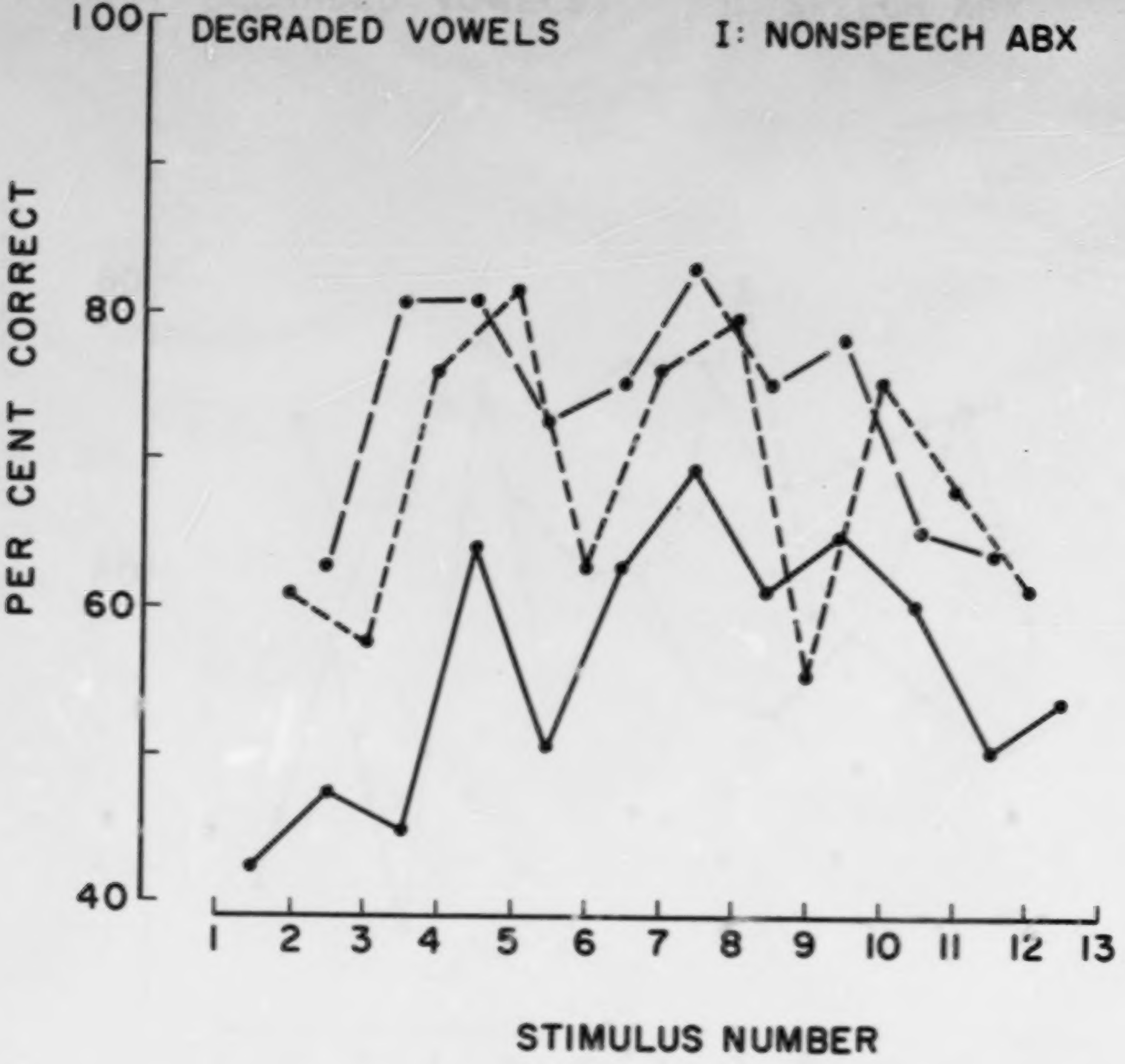

DEGRADED VOWELS
I: NONSPEECH ABX
100
80
60
40
PER CENT CORRECT
1
2
3
4
5
6
7
8
9
10
11
12
13
STIMULUS NUMBER

Fig. 4

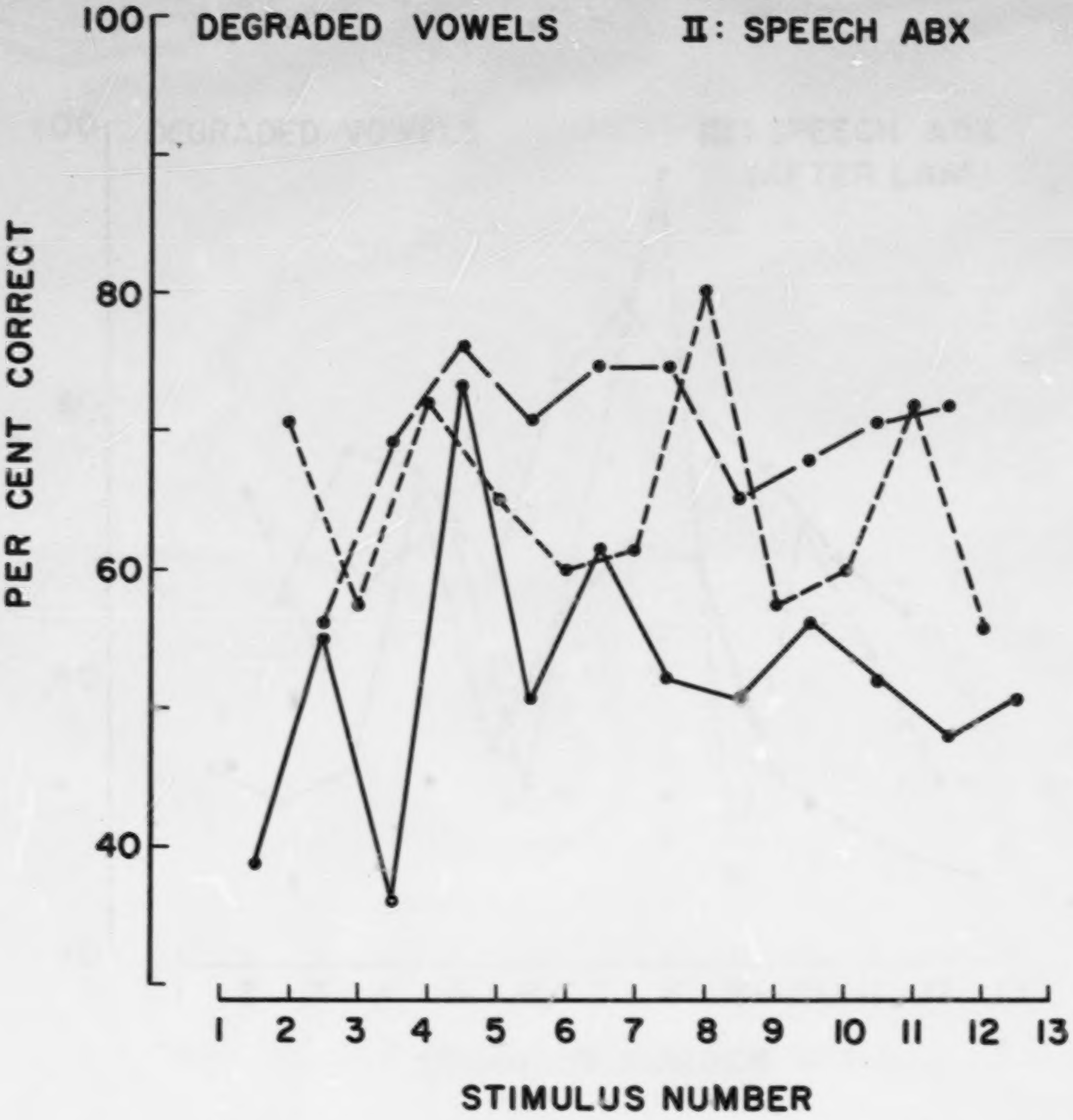

DEGRADED VOWELS
II: SPEECH ABX
100
80
60
40
PER CENT CORRECT
1
2
3
4
5
6
7
8
9
10
11
12
13
STIMULUS NUMBER

Fig. 5

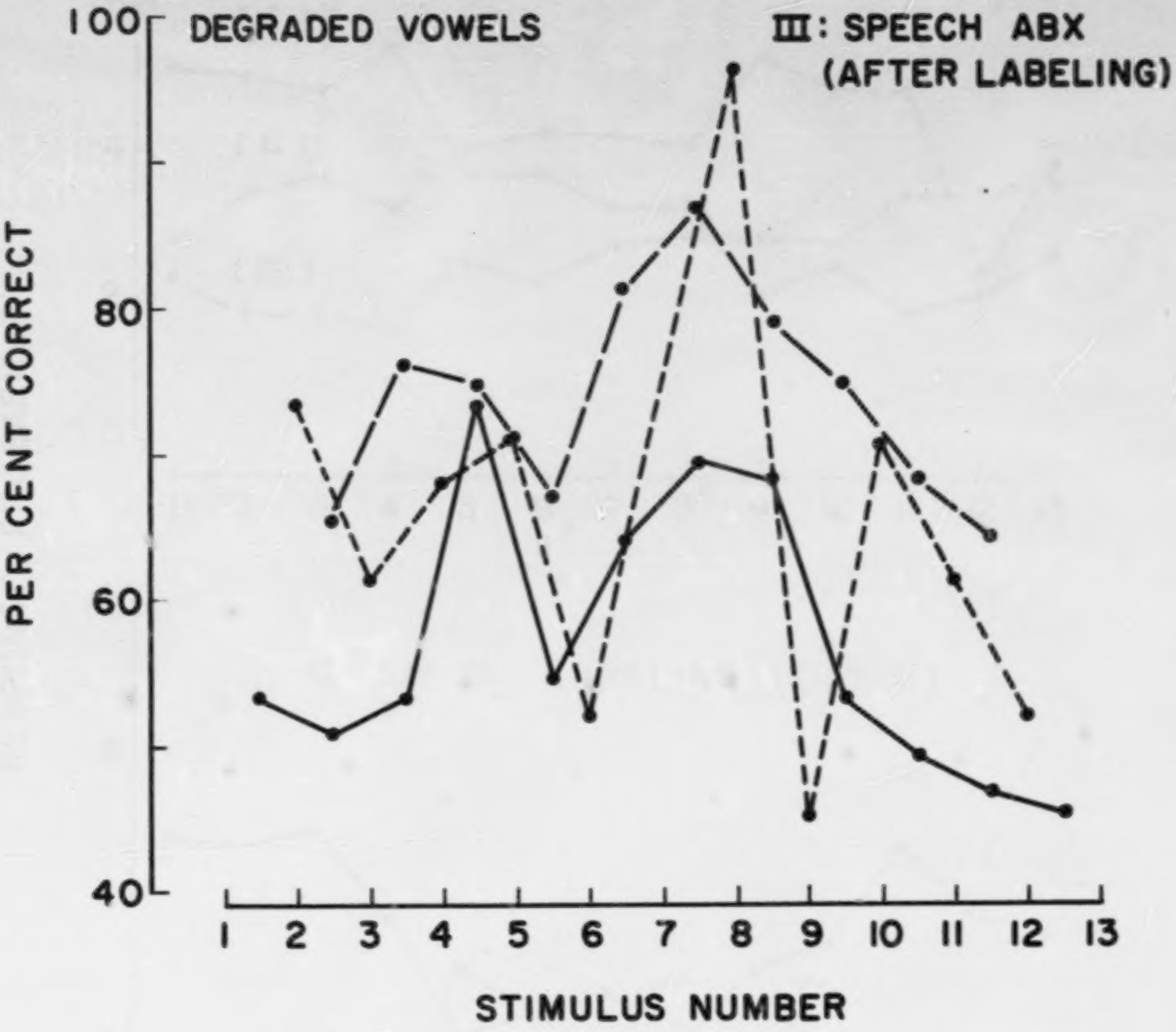
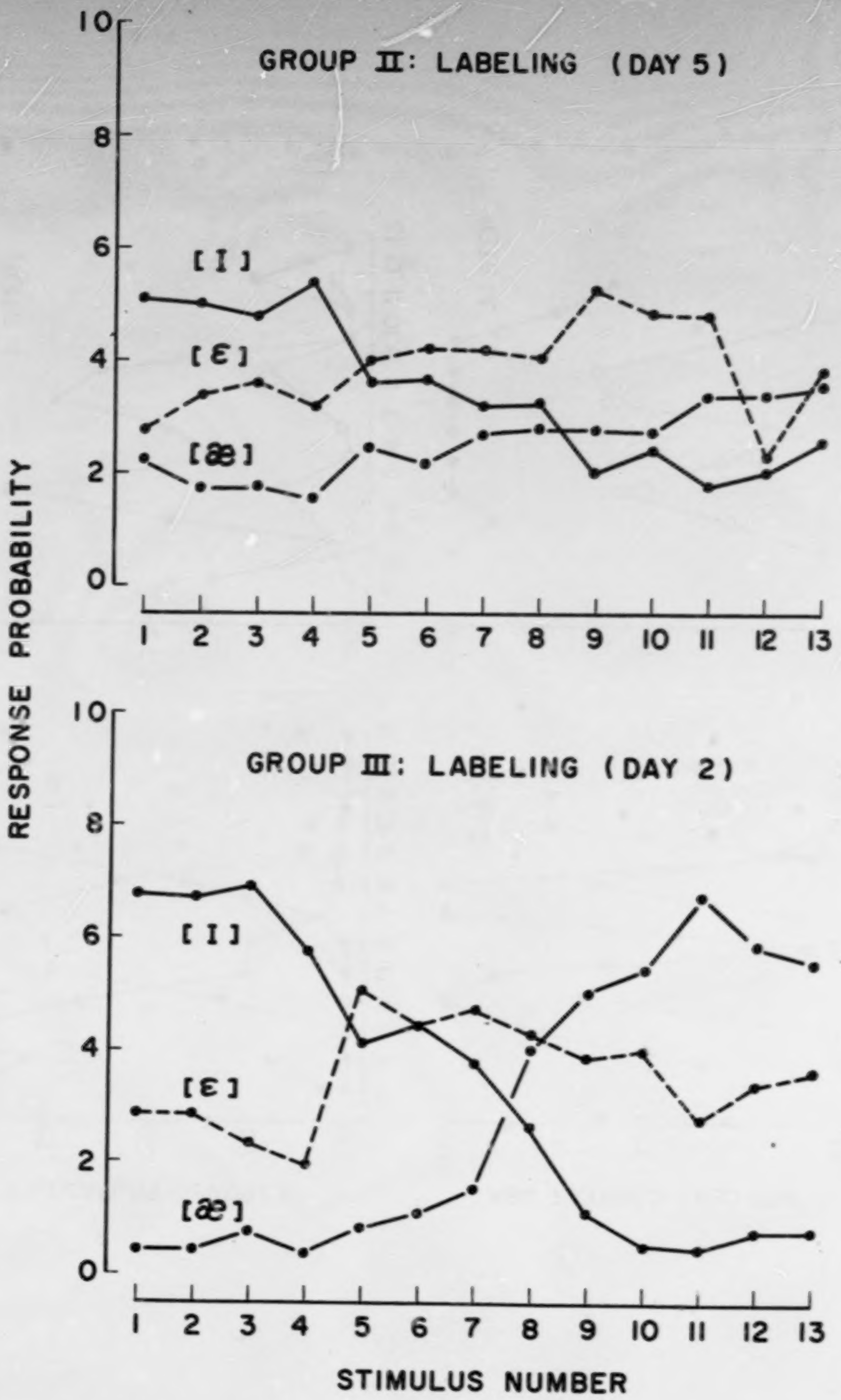

DEGRADED VOWELS
III: SPEECH ABX
(AFTER LABELING)
100
80
60
40
PER CENT CORRECT
1
2
3
4
5
6
7
8
9
10
11
12
13
STIMULUS NUMBER

Fig. 6


GROUP II: LABELING (DAY 5)

GROUP III: LABELING (DAY 2)

Fig. 7

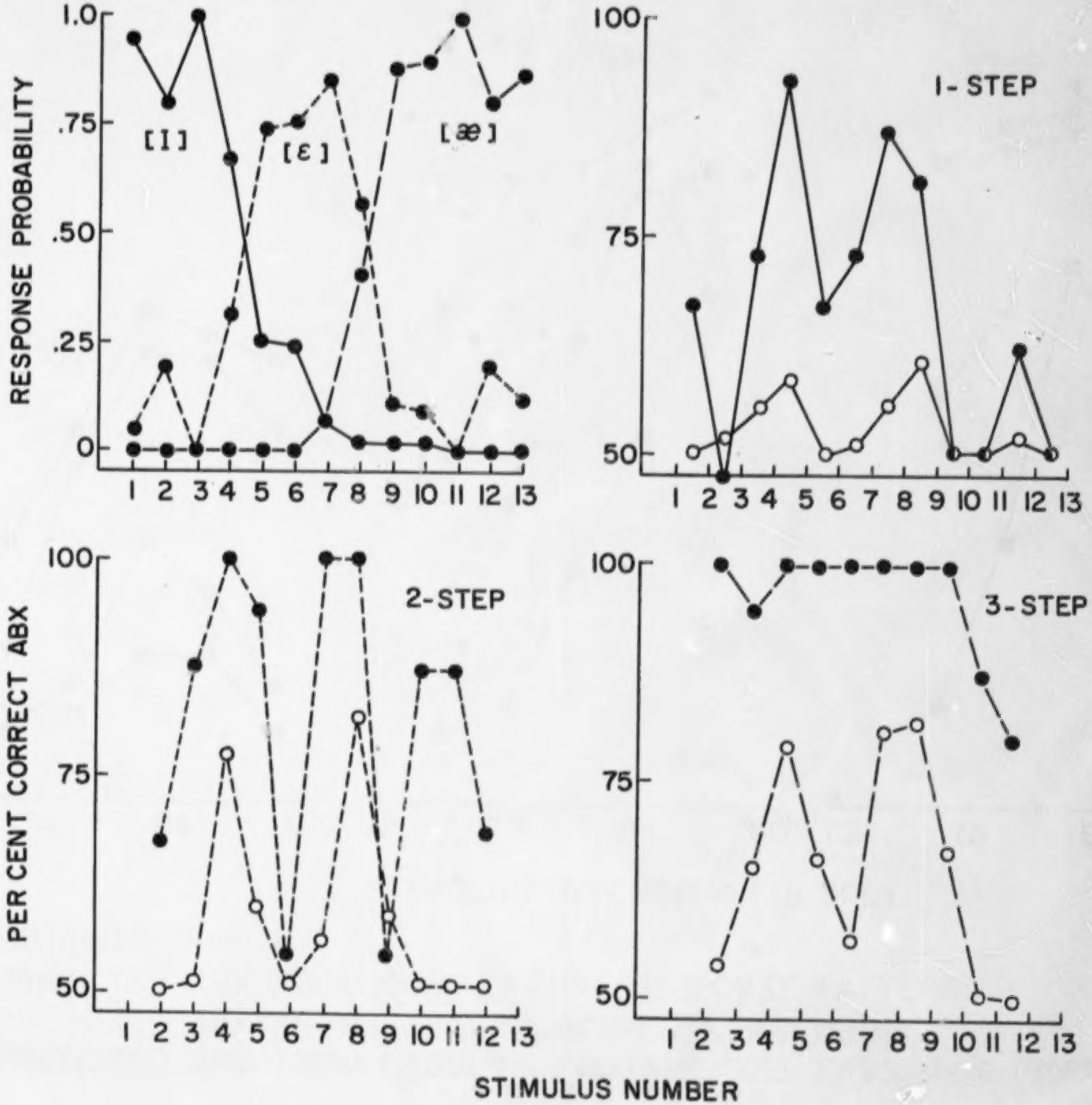

RESPONSE PROBABILITY
1.0
.75
.50
.25
0
[I]
[ε]
[æ]
1 2 3 4 5 6 7 8 9 10 11 12 13
100
75
50
1- STEP
1 2 3 4 5 6 7 8 9 10 11 12 13
PER CENT CORRECT ABX
100
75
50
2-STEP
1 2 3 4 5 6 7 8 9 10 11 12 13
100
75
50
3-STEP
1 2 3 4 5 6 7 8 9 10 11 12 13
STIMULUS NUMBER

Fig. 8

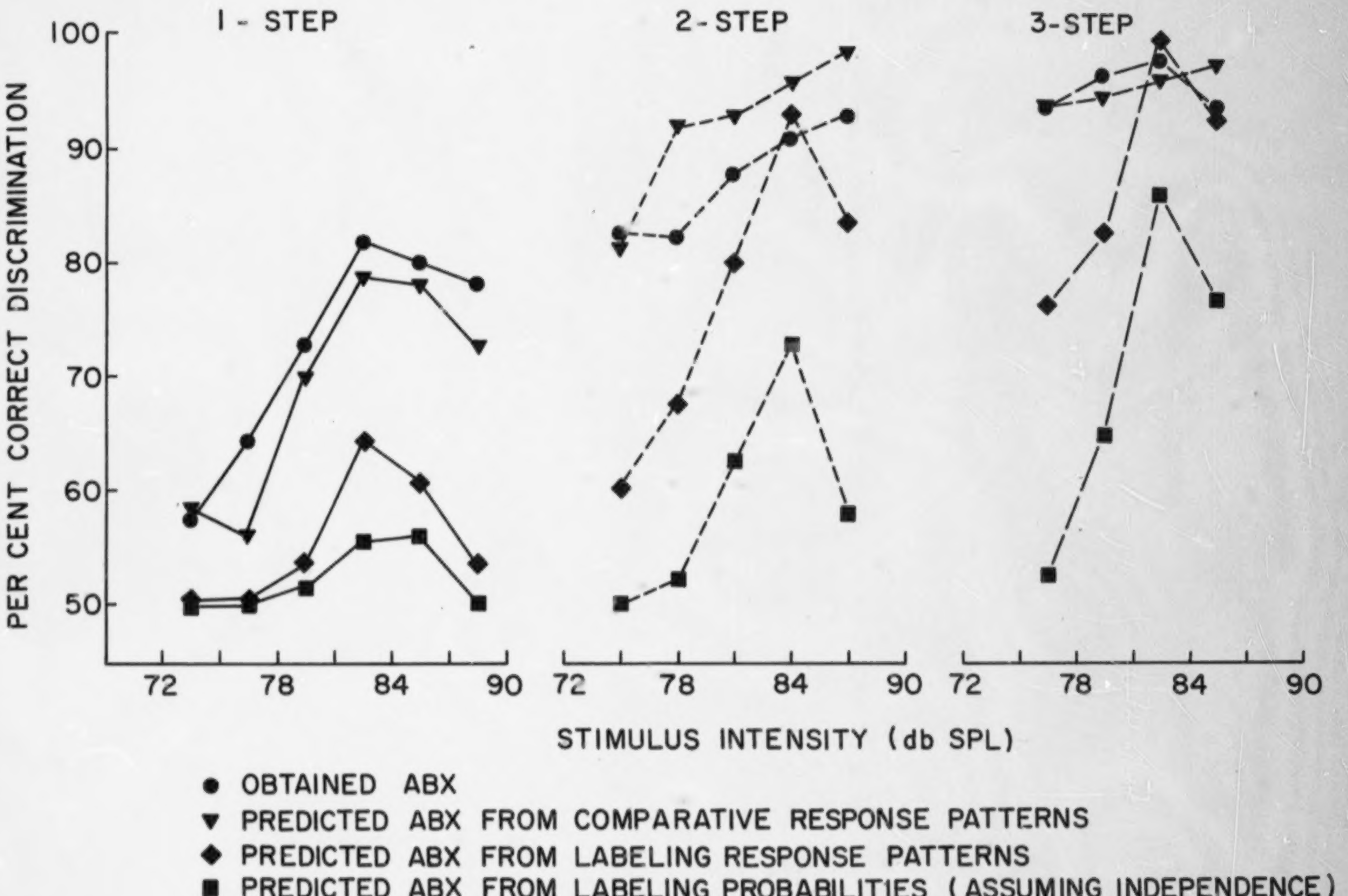

1 - STEP
2 - STEP
3 - STEP
PER CENT CORRECT DISCRIMINATION
100
90
80
70
60
50
72
78
84
90
72
78
84
90
72
78
84
90
STIMULUS INTENSITY (db SPL)
● OBTAINED ABX
▼ PREDICTED ABX FROM COMPARATIVE RESPONSE PATTERNS
◆ PREDICTED ABX FROM LABELING RESPONSE PATTERNS
■ PREDICTED ABX FROM LABELING PROBABILITIES (ASSUMING INDEPENDENCE)

Fig. 9

IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR GRAPHEMES AND LINES

J. L. Kopp and D. V. Cross
The University of Michigan

Identification and Discrimination Functions for Graphemes and Lines

J. L. Kopp and D. V. Cross

Behavior Analysis Laboratory

Several recent studies performed in this laboratory (Cross & Lane, 1962; Cross, Lane & Sheppard, 1964; Lane, 1964; Lane & Kopp, 1964) have measured generalization of two mutually incompatible responses to stimuli on a continuum intermediate to their controlling stimuli. The generalization gradients obtained in this way divide the continuum into two discrete response classes. Discrimination (ABX) functions, obtained after the generalization test, indicate that discrimination is better at stimulus values which lie near the boundary of the two response classes, i.e., near the point where the two responses are equiprobable.

The present study examines the relation between concurrent generalization gradients and the corresponding discrimination functions for visual stimuli of two kinds: (a) variants of the alphabetical letters E and F, for which the baseline was varied in length so as to define a continuum anchored at one end by the letter E and at the other by the letter F; and (b) horizontal lines corresponding in length to the baselines of the letters used above.

In view of our related findings for other stimulus continua, we may anticipate that, in the case of the alphabetical letters, categorical responding will occur as a result of individual histories of discriminative reinforcement simply by instructing subjects to identify the stimuli as either one letter or the other; discrimination functions taken afterward will show the characteristic peaks at the boundary of the two response classes. In the case of lengths of lines, categorical responding should follow after conditioning two arbitrary responses to the extreme stimulus values. Discrimination functions taken afterward should also show peaks at the point of response equiprobability. In either case, the prediction of discrimination enhancement is predicated on the prior establishment of categorical labelling behavior--either by the language community alone (case a) or with the intervention of the experimenter (case b). The present study was designed to examine these predictions, as they apply to linguistic and nonlinguistic visual continua, and to determine if discrimination enhancement would be observed without the prior establishment of differential responding.

Method

Four experimental conditions were used and ten Ss assigned to each condition as follows:

Group A - Generalization testing of the letters E and F followed by one-step discrimination testing, in which ABX triads involved comparisons of pairs of adjacent stimuli on the continuum.

Group B - Generalization testing of the letters E and F followed by two-step discrimination testing, in which ABX triads involved comparisons of pairs of stimuli separated by two steps on the continuum.

Group C - Discriminative training with lengths of lines, followed by generalization testing and then by one-step discrimination testing.

Group D - Discrimination testing with lengths of lines, followed by discriminative training and then generalization testing.

The alphabetical stimuli were letters .797 in. high by .578 in. wide, with baselines of seven different lengths (.25 to .578 in. in logarithmic steps). The line stimuli consisted of just the baselines of the letters. The stimuli were formed by mounting strips of black vinyl tape .0625 in. wide on white fiber board discs. The discs were then covered with semi-opaque tracing paper (to obscure minute defects in the letters)and placed beneath a 1/16 in. Plexiglas sheet. The stimuli were viewed from a distance of about 3 ft., at eye level, through a 2 in. aperture, which was positioned 2½ in. behind a solenoid-operated annular tachistoscope (Burke & James, Inc.). The surround for the 2 in. aperture was painted flat black and illuminated by two miniature neon bulbs.

The stimulus protocols were as follows:

a) Discriminative training. The extreme stimuli from the series were presented 25 times each in a quasi-random order. (Total = 50)

b) Generalization testing. All of the stimuli were presented randomized in blocks of 7 with 10 blocks to the series. (Total = 70)

c) One-step ABX discrimination test. The 24 possible combinations were presented three times each in quasi-random order. (Total = 72)

d) Two-step ABX discrimination test. The 20 possible combinations were presented four times each in quasi-random order. (Total = 80)

During the discrimination training with lengths of lines, the subjects were instructed as follows:

> "In this phase of the experiment you can earn money by labelling lengths of lines. As the stimuli appear behind the glass on the display unit, you are to call out 'A' or 'B' whichever you think appropriate. Each time you respond appropriately, a penny will fall into the tray in front of you. The amount of money you will be paid for your time here will be partly determined by the number of pennies you earn."

In order to assess generalization for lengths of lines, these instructions were then given:

> "In this phase of the experiment, you are to continue responding as before. However, you will not receive pennies from the apparatus. Instead, a record of your correct responses will be kept and you will be paid accordingly at the end of the session."

The generalization of the responses to the alphabetical letters was measured after administering these instructions:

> "In this phase of the experiment you will be shown variations of two letters of the alphabet. The stimuli will appear behind the glass on the display unit in front of you. You are to respond after each stimulus presentation by calling out the name of the letter. If you are uncertain, guess. Be sure to respond after each stimulus is presented."

The instructions for discrimination testing were as follows:

> "In this experiment you will be shown various stimuli in sequence, three stimuli per sequence. The stimuli will appear behind the glass on the display unit in front of you. The third stimulus in each sequence will always be identical to one of the first two stimuli. After each sequence, you are to say 'A' if the third stimulus is like the first stimulus, or 'B' if the third stimulus is like the second. If you are uncertain, guess. Be sure to respond after each three-stimulus sequence."

After the subjects in group B had completed the experiment, psychophysical data were taken as part of another study (Kopp & Cross, 1964).

Results and Discussion

Figure 1 shows the identification and discrimination functions for each of the four groups. Discriminations among the one-step comparisons are fairly uniform at about the 70 percent correct level. For the two-step ABX task, discrimination is closer to 100 percent and again rather uniform across the continuum. It can be seen that there is no clear-cut trend for discrimination to be better at the point of equiprobability for the labelling responses.

The identification functions (generalization gradients) clearly reflect categorical responding for all groups. The point of equal response probability is very similar for the groups who were presented alphabetical letters (A and B). The other two groups yielded gradients whose points of equiprobability were clearly different. The procedure of averaging individual data for identification and discrimination may mask a correlation between the two if the points of response equiprobability or discrimination enhancement for individual subjects are dispersed along the stimulus continuum (Lane, 1964). Examination of individual data for each of the four groups, shown in Figs. 2-5, reveals that a consistent relation between identification and discrimination was not obtained.

Since discrimination enhancement has regularly been found to accompany categorical labelling in prior research with linguistic and nonlinguistic continua, the failure to find this relation in the present study may be attributable to a procedural artifact. As explicated by another article in this report (Cross & Lane, 1964) the spacing of the stimuli sampled from the continuum under study largely determines the sensitivity of the discrimination test and hence the detection of discrimination enhancement when present. Apparently it is optimal to use an array of stimuli that are closely packed, relative to the resolving power of the sensory modality involved, so that discrimination accuracy is near chance levels within labelling categories. When this condition does not obtain, as in the study of the vowels by Fry _et al_. (1962) and in the present experiment (cf. Fig. 1), and the test stimuli are too widely spaced, gross psychophysical differences may overwhelm the effects due to the prior establishment of categorical identification.

References

Cross, D. V., & Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency and topography in auditory generalization. *J. exp. anal. Behav.*, 1962, 5, 487-496.

Cross, D. V., & Lane, H. L. An analysis of the relations between identification and discrimination functions for speech and nonspeech continua. Unpublished report, 05613-3-P, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, 1964.

Cross, D. V., Lane, H. L., & Sheppard, W. Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. *J. exp. Psychol.*, 1964, in press.

Kopp, J. L., & Cross, D. V. A relation between stimulus generalization and magnitude estimation. Unpublished report, 05613-3-P, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, 1964.

Lane, H. L., & Kopp, J. L. The effects of response dependent and independent reinforcement in extending stimulus control. *Psychol. Rec.*, 1964, 14, 81-87.

Lane, H. L. The motor theory of speech perception: a critical review. *Psychol. Rev.*, 1964, in press.

Figure Captions

Fig. 1. Identification and discrimination functions for the grapheme contrast E-F and for lengths of lines. Groups A and B received generalization and discrimination testing with variants of the graphemes separated by one or by two steps, respectively, along the stimulus continuum. Group C received discriminative training with the extreme stimuli of the series, followed by generalization and discrimination testing. Group D received discrimination testing first, then training and generalization testing.

Fig. 2. Identification and discrimination functions for individual Ss in group A. Each datum point represents the mean of 10 or of 12 determinations, respectively.

Fig. 3. Identification and discrimination functions for individual Ss in Group B. Each datum point represents the mean of 10 or of 16 determinations, respectively.

Fig. 4. Identification and discrimination functions for individual Ss in group C. Each datum point represents the mean of 10 or of 12 determinations, respectively.

Fig. 5. Identification and discrimination functions for individual Ss in group D. Each datum point represents the mean of 10 or of 12 determinations, respectively.

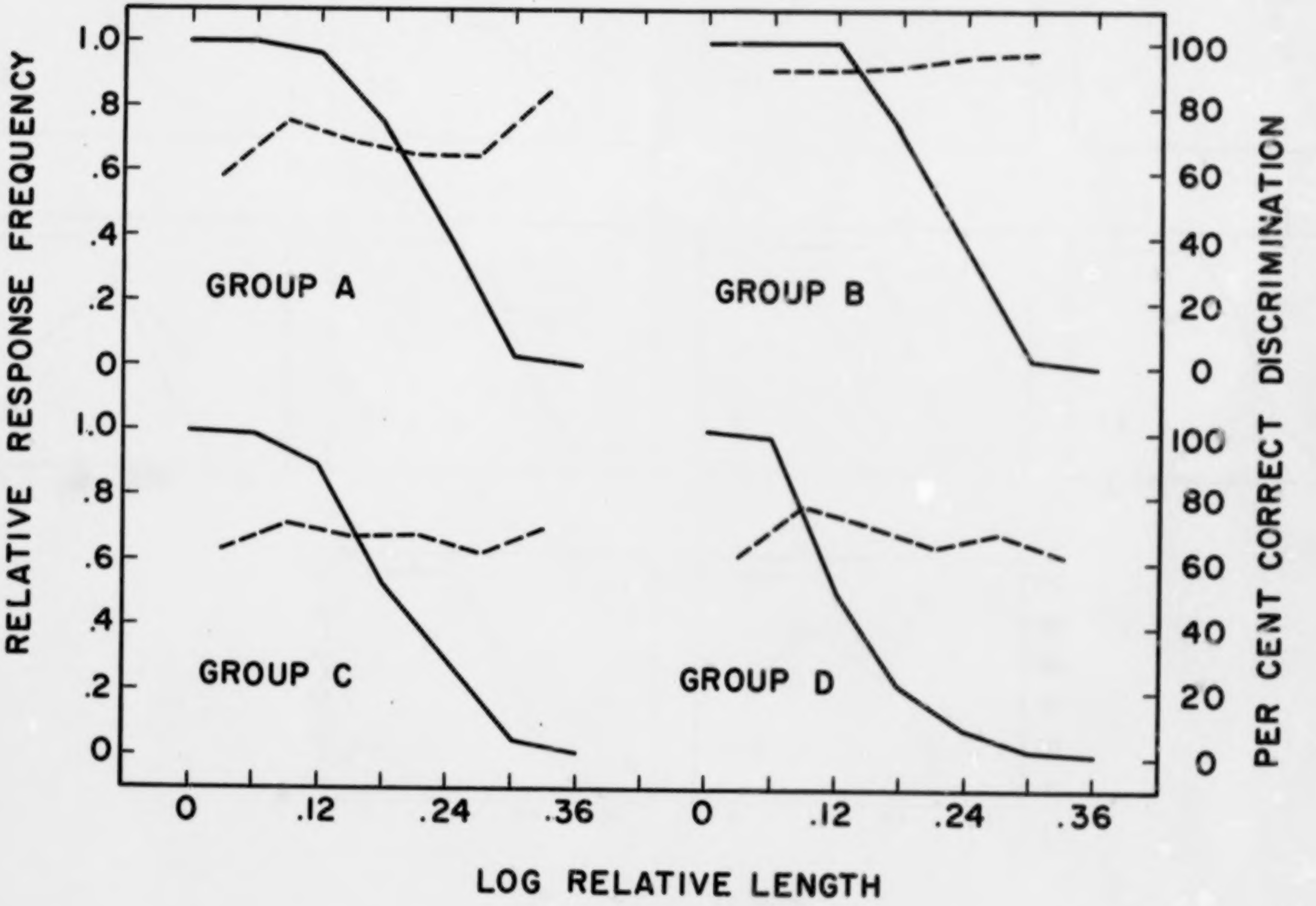
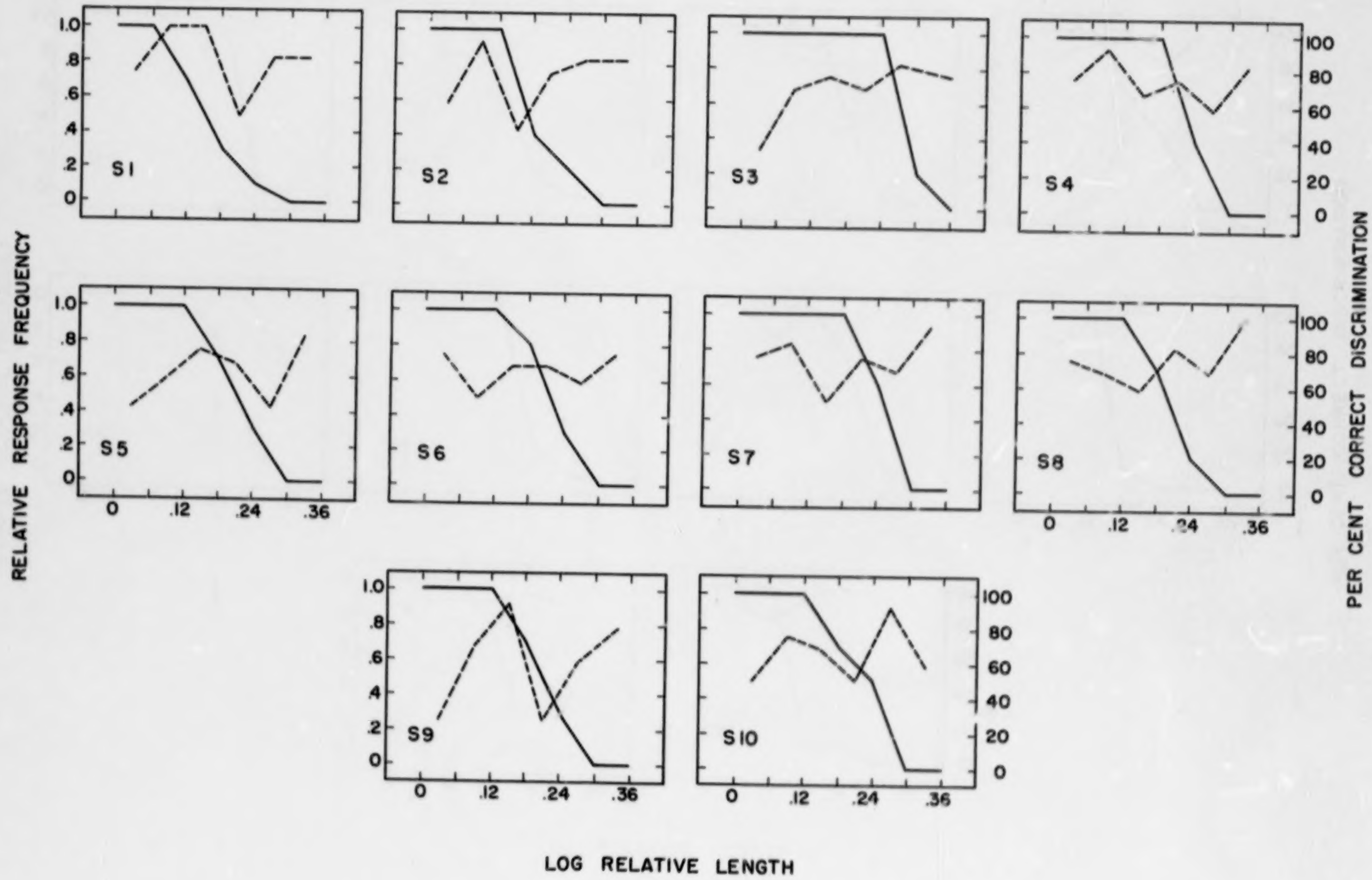
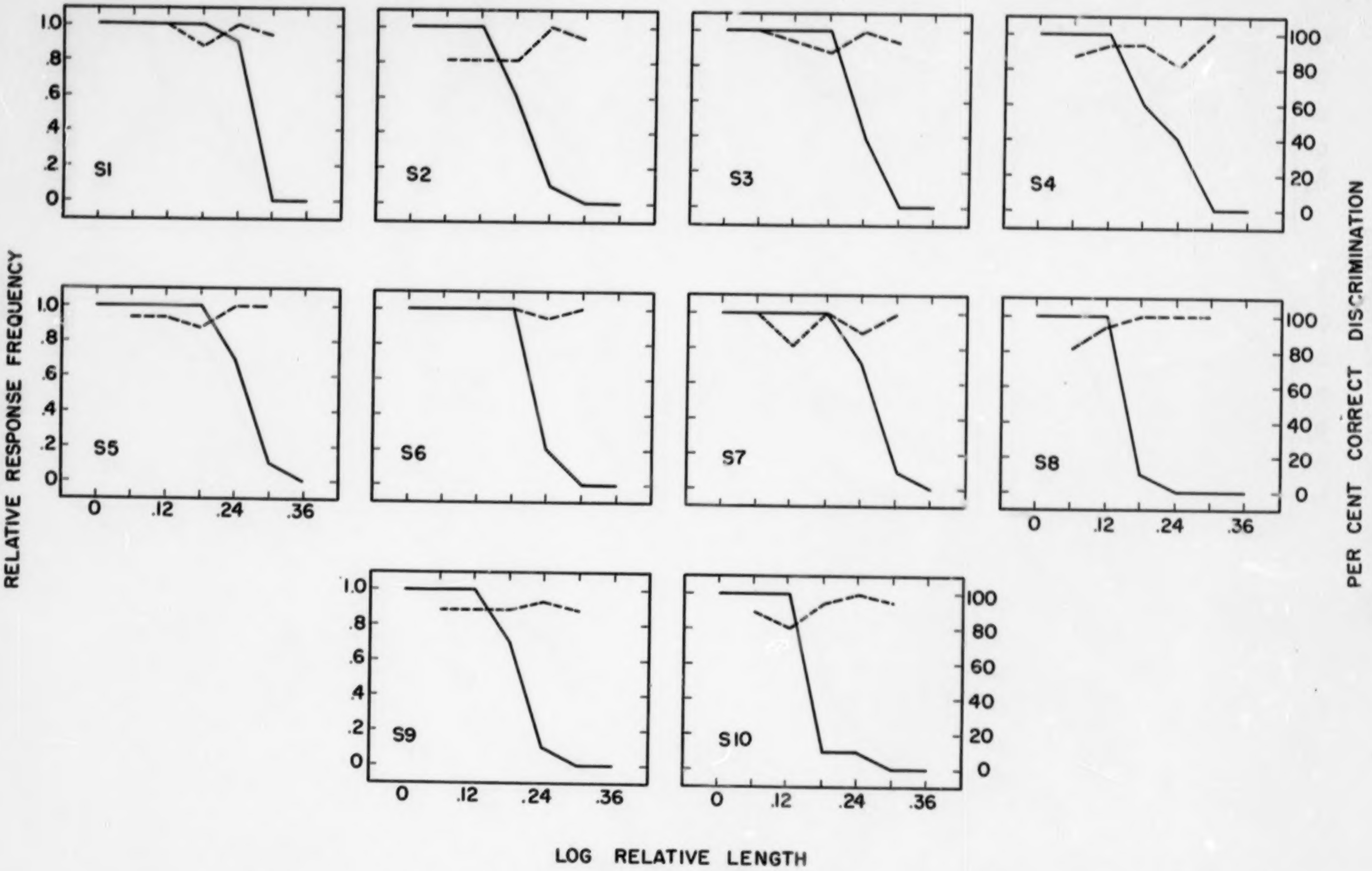

1.0
.8
.6
.4
.2
0
1.0
.8
.6
.4
.2
0
RELATIVE RESPONSE FREQUENCY
GROUP A
GROUP B
GROUP C
GROUP D
100
80
60
40
20
0
100
80
60
40
20
0
PER CENT CORRECT DISCRIMINATION
0
.12
.24
.36
0
.12
.24
.36
LOG RELATIVE LENGTH

Fig. 1


RELATIVE RESPONSE FREQUENCY
PER CENT CORRECT DISCRIMINATION
LOG RELATIVE LENGTH
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
1.0
.8
.6
.4
.2
0
100
80
60
40
20
0
0
.12
.24
.36

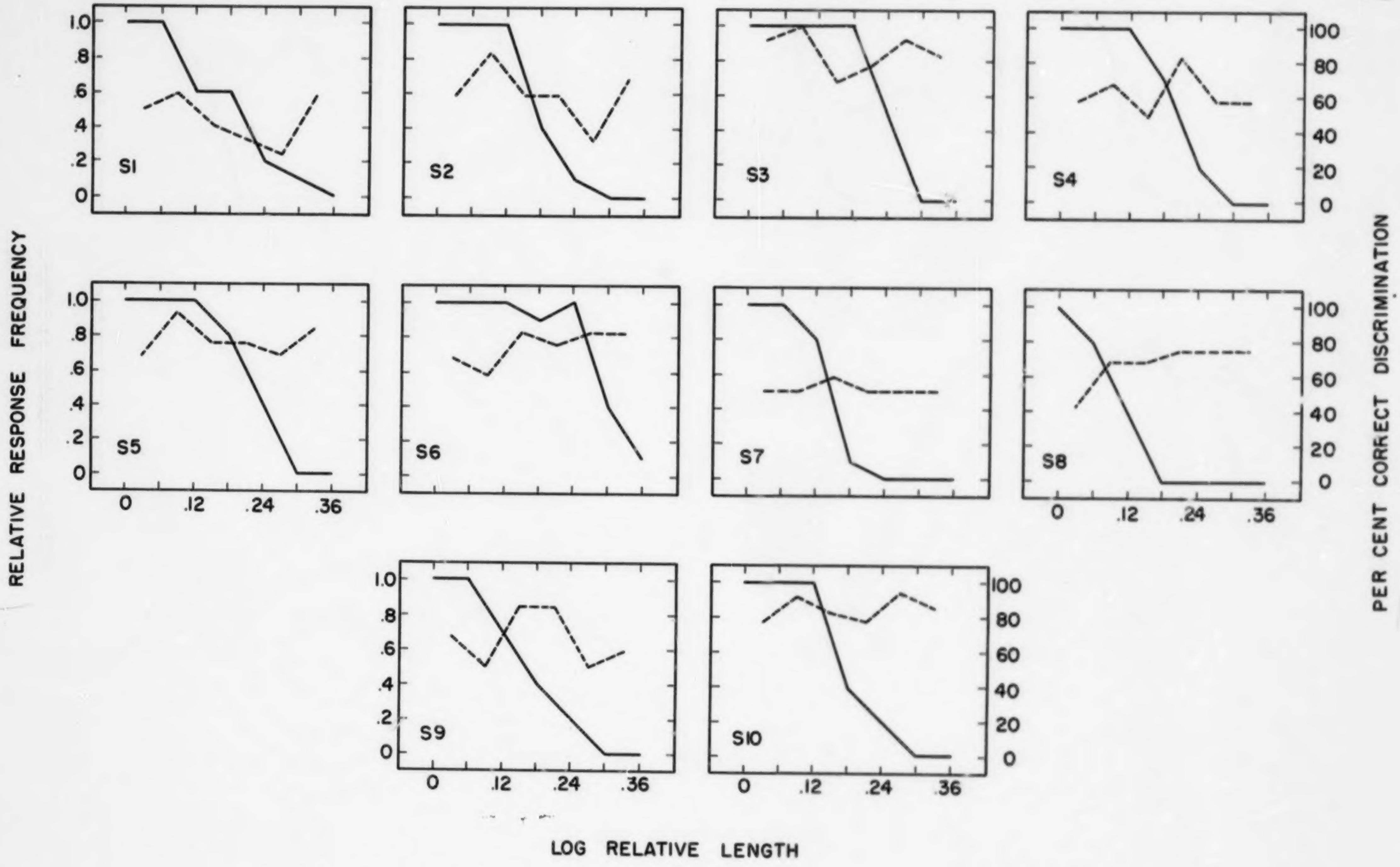
GROUP A

Fig. 2


RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
PER CENT CORRECT DISCRIMINATION
100
80
60
40
20
0
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
0
.12
.24
.36
LOG RELATIVE LENGTH

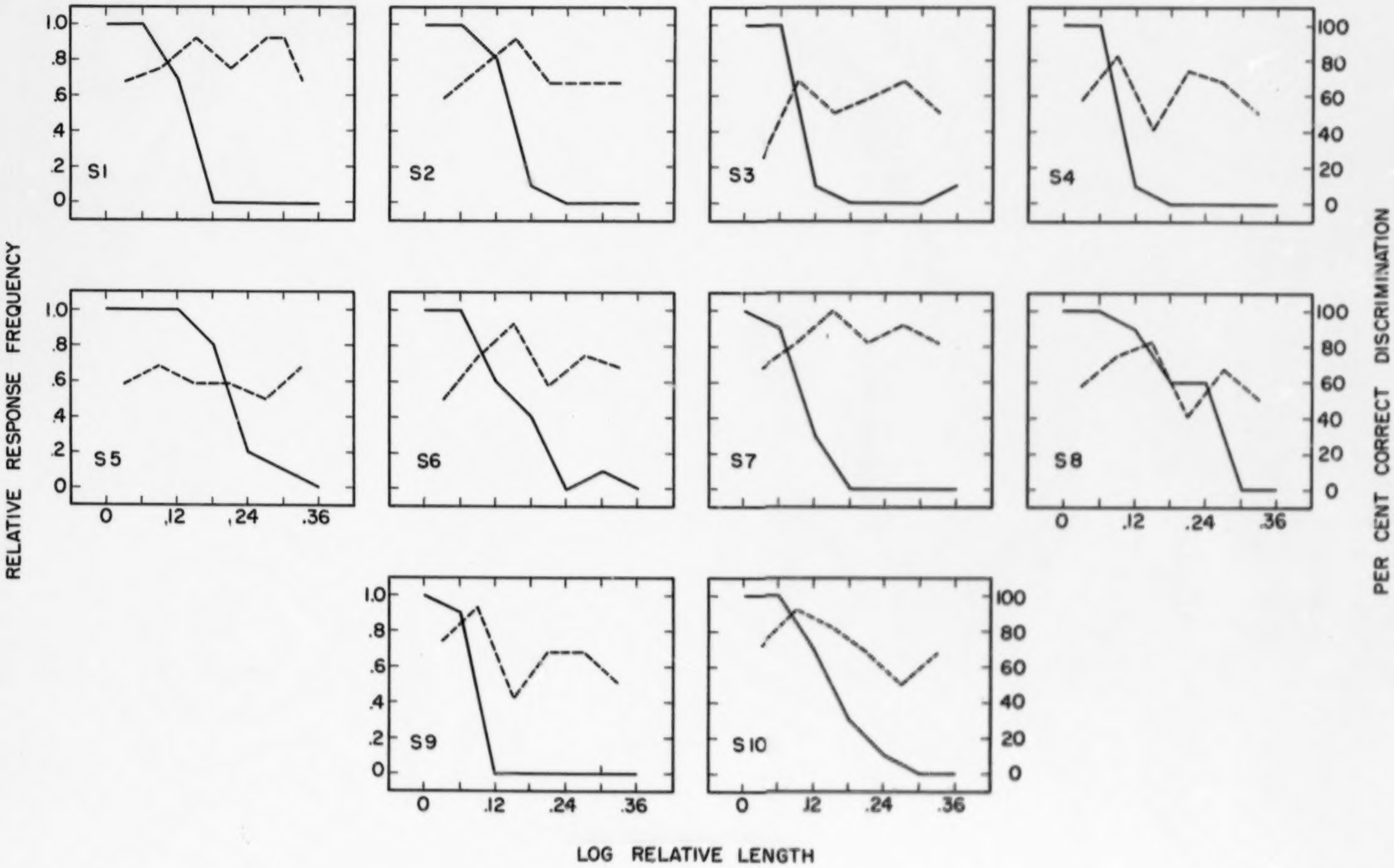
GROUP B

Fig. 3


S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
PER CENT CORRECT DISCRIMINATION
100
80
60
40
20
0
0
.12
.24
.36
LOG RELATIVE LENGTH

GROUP C

Fig. 4


S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
PER CENT CORRECT DISCRIMINATION
100
80
60
40
20
0
.12
.24
.36
LOG RELATIVE LENGTH

GROUP D

Fig. 5

A RELATION BETWEEN STIMULUS GENERALIZATION AND MAGNITUDE ESTIMATION

J. L. Kopp and D. V. Cross
The University of Michigan

A Relation between Stimulus Generalization and Magnitude Estimation

J. L. Kopp and D. V. Cross

Behavior Analysis Laboratory

In an earlier study performed at this laboratory (Cross and Lane, 1962), two discriminative stimuli from an intensity continuum were conditioned to two different responses; stimulus generalization was measured concurrently for the two responses by presenting stimulus intensities intermediate between the training stimuli. It was found that the complementary generalization gradients obtained in this way partitioned the stimulus continuum into two discrete response classes, bounded at the point where the two responses are equally probable. A second study (Lane and Kopp, 1964) replicated these results. It wasobserved in both of these studies that the point on the continuum corresponding to equi-probable responses also corresponded to the midpoint of the sone scale values for the intensities involved. A pilot study was then carried out in similar fashion, using pure tone duration and electric shock for the stimulus continua. The results confirmed the conclusion that the point of equal response probability tends to co-vary with a psychophysical midpoint predicted from data obtained in other laboratories by direct scaling methods. However, the generalization midpoint tended to be about 0.5 db below that predicted from the psychophysical scale (where $N_{db} = 10 \log x_1/xo$). This finding is similar to the results of other studies relating bisection to direct scaling data (Newman, Volkmann, and Stevens, 1937).

If a systematic relation can be demonstrated between scales of apparent magnitude based on stimulus generalization and those based on direct estimation, psychophysical methods would be not only better understood but also applicable, for the first time, to the measurement of sensory dynamics in nonverbal human and infra-human organisms. The present study examines the relation between magnitude estimation and stimulus generalization with two intensive continua: lengths of lines and intensities of pure tones.

Method

In experiment I the stimuli were seven lengths of lines varying from .250 to .578 inches in equal logarithmic steps. Ten paid subjects (Ss) were individually conditioned to give one of two responses to each of the two extreme stimulus values. Generalization was measured during a subsequent phase in which S was instructed to emit one of the two responses to each of the seven stimulus values, presented randomly in ten blocks of seven stimuli each. A discrimination task followed the generalization phase as part of another experiment. The stimulus protocols, instructions, and apparatus used for these phases of the experiment are described elsewhere in this report (Kopp & Cross, 1964). In a final phase each S gave magnitude estimations of the lengths of lines after receiving these instructions:

> "In this phase of the experiment you are asked to give numerical estimates of lengths of lines. The first line you will see will be the standard which you are to call 'ten.' When you see it say 'ten.' Then, when you see the second line, tell me the numerical value of its length and so on throughout the series. Always assign numbers to the stimuli in the same proportion to ten as their length is to the standard. For example, if the second stimulus is twice as long as the standard, call it 'twenty,' if it is half as long, call it 'five,' one

third as long, '3.3,' etc. Are there any questions? All right let's begin. Remember, the first stimulus is the standard and you are to call it 'ten.'"

Questions were answered by repeating the instructions. The same stimulus series was used in the magnitude estimation phase as was used in the generalization testing phase.

In experiment II the stimuli were eleven 1,000 cps tones varying in intensity from 50 db to 100 db (SPL). The stimuli were recorded on magnetic tape and presented to S through earphones (Grason-Stadler TDH-39 w/neoprene cushions.) The stimulus duration was one sec. with a 5 sec. inter-stimulus interval. An electronic switch (Grason-Stadler 829C), with a 25 msec. rise and decay time, cut off the input to the earphones between stimuli thereby preventing print-through and other noises from reaching S. A band-pass filter was interposed between the switch and the phones to further minimize noise. Each S was tested in a sound-attenuating chamber and instructions were given over an intercom. During the conditioning phase, the S was given pennies for pushing one or the other of two buttons on a panel in response to one of the two extreme stimulus values. Conditioning was continued to a criterion of 20 correct responses. Generalization was measured during the ensuing (extinction) phase in which the stimuli were presented randomly in blocks of eleven for a total of 110 presentations. After generalization testing, S gave magnitude estimations of the same stimulus series. Instructions for the three phases of this experiment were similar to those used in experiment I. Ten Ss served under the conditions just described and ten additional Ss gave the magnitude estimations prior to conditioning.

Experiment III was exactly like experiment II except that the 20 Ss were each conditioned to push one of three buttons on a panel in response to either the 50 db, the 75 db or the 100 db (SPL) tones during the conditioning phase.

Results

Experiment I

Figure 1 shows the generalization gradient and Fig. 2 the magnitude estimation function for all ten Ss combined. The generalization gradient is orderly and of the form reported in the earlier studies (e.g., Cross & Lane, 1962). The gradient for the alternate response is complementary and is not shown.

A least squares fit to the average magnitude estimations yields a line whose slope is 1.26, which is fairly close to the slope of 1.1 considered characteristic for judgments on this continuum (Stevens, 1957). The slopes for individuals ranged from .9 to 1.74. It is possible to examine the relation between the bisection point of the stimulus continuum according to the generalization gradient and that predicted from direct estimation in the following way. A point is plotted in Fig. 2 with an ordinate value equal to one-half the range of the magnitude estimations (encompassed by the line of best fit) and an abscissa value corresponding to the point of response equiprobability in Fig. 1. This point is then located .036 units on the X-axis below the psychological midpoint according to magnitude estimation. The displacement of individual points varies from -.119 to +.068 relative log units.

A category scale of lengths of lines is constructed from the relative response frequencies shown in Fig. 1 by multiplying the relative frequency of R_1 responses at each stimulus value by unity and that of R_2 responses by two and then adding these weighted frequencies for each stimulus. The category rating thus assigned to each stimulus is plotted as a function of its average magnitude estimation in Fig. 3. The function relating the derived category scale to the direct estimation scale is concave downward.

Experiment II

Generalization gradients of pure tone intensity are shown in Figs. 4 and 5 for the two groups of Ss. It can be seen that the two gradients would be nearly superimposed if plotted on the same graph; they both partition the stimulus continuum into two classes.

Figures 6 and 7 show the magnitude estimation functions for the two groups. The slope of the function for the group which gave magnitude estimations in the first phase of the experiment is .49 with individuals ranging from .37 to .71, whereas that for the group in which magnitude estimations followed the generalization phase is .28 with individual slopes ranging from .17 to .44.

The points of equal response probability obtained from the group generalization gradients are plotted on the graphs of the magnitude estimation function at abscissa values corresponding to one half of the range subtended by this function. This point is located 3.52 db below the magnitude estimation midpoint for the group in which magnitude estimations were taken first and .51 db above this midpoint for the group in which magnitude estimations constituted the final phase. Individuals ranged from -13.16 db to +2.99 db for the former group and from -11.70 to + 9.35 for the latter.

A category scale of pure tone intensity is constructed from the relative response frequencies shown for the two groups in Figs. 4 and 5 according to the method described above for scaling lengths of lines. The two derived category scales are plotted as a function of the corresponding direct estimation scales in Figs. 8 and 9. The concave downward function is again obtained--most clearly in the case of those Ss who gave magnitude estimation before discriminative conditioning (Fig. 8).

Experiment III

Generalization gradients of pure tone intensity are shown in Figs. 10 and 11 and the corresponding magnitude estimation functions in Figs. 12 and 13 for the two groups of Ss respectively.

The generalization functions effectively partition the continuum into three discrete classes corresponding to the three responses reinforced during the conditioning phase of the experiment.

For the group in which magnitude estimation preceded the conditioning and generalization phases, the slope of the magnitude estimation function is .40, with individual data ranging from .25 to .61. In the other group, the slope is decidedly less, .26, with individuals ranging from .21 to .34.

For the former group a comparison of the points of equiprobability and the points on the magnitude estimation function corresponding to 1/4 and 3/4 of the total range of abscissa values shows the lower points to differ by 7.67 db and the upper points to differ by 5.51 db with the points based on the generalization gradients being the lesser in both cases. Individuals ranged from -13.01 db to -1.36 for the lower points and -8.68 db to -0.42 db for the upper. The same data for the group that gave magnitude estimates in the last phase of the experiment show the lower (generalization) point to be -1.32 db and the upper -2.12 db from the magnitude estimation function, with individuals ranging from -5.92 db to +5.41 db for the lower point and -7.77 db to +4.05 db for the upper point.

The relations between the derived category scales and the direct estimation scales may be examined for the two groups in Figs. 14 and 15. Once again, the finding is that the category scale is concave downward against the magnitude scale if the latter is uncontaminated by the effects of discrimination training (Fig. 14).

Discussion

The findings of the three experiments confirm and extend prior research in revealing orderly generalization gradients, following discrimination training with both two and three concurrent responses, that effectively partition the stimulus continuum into discrete classes, bounded by the points where responses are equiprobable.

The points of response equiprobability may be treated as psychophysical equisection points because (1) the corresponding stimulus values covary with (are typically a few decibels less than) what would be predicted from direct scaling data, and (2) the functions relating category scales, derived from the generalization gradients, to magnitude estimates yield the characteristic curvilinear relationship which is concave downward.

Although the generalization gradients were not significantly affected when the conditioning phase was preceded by magnitude estimation, the magnitude estimation functions were significantly flattened, and the variability of the slopes of individual functions significantly decreased, when magnitude estimation was conducted after discrimination training and generalization testing. Under the latter conditions, the Ss tended to use fewer numbers and make less extreme judgments. This outcome probably results from the high degree of categorical stimulus control established by discrimination training.

References

Cross, D. V., & Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. J. exp. anal. Behav., 1962, 5, 487-496.

Kopp, J. L., & Cross, D. V. Identification and discrimination functions for graphemes and lines. Unpublished report 05613-3-P, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, 1964.

Lane, H. L., & Kopp, J. L. The effects of response dependent and independent reinforcement in extending stimulus control. Psychol. Rec., 1964, 14, 81-87.

Newman, E. B., Volkman, J., & Stevens, S. S. On the method of bisection and its relation to a loudness scale. Amer. J. Psychol., 1937, 49, 134-147.

Stevens, S. S. On the psychophysical law. Psychol. Rev., 1957, 64, 153-181.

Figure Captions

Fig. 1. Stimulus generalization to lengths of lines following discrimination training. Each extreme stimulus was correlated with reinforcement for one of two responses. Since the two gradients are complementary only the R_1 gradient is plotted. Each point represents 10 determinations with each of 10 Ss.

Fig. 2. Magnitude estimations of the lengths of lines. The average estimates of individual Ss were converted to relative units and the geometric means for the pooled averages were plotted. Each point represents 10 estimates from each of 10 subjects.

Fig. 3. Relationship between magnitude estimations and category estimations constructed from the two-response generalization gradients for lengths of lines. The relative frequency of R_1 at each stimulus value was multiplied by unity and that of R_2 was multiplied by two. The arithmetic sum of these weighted frequencies is taken as the category assignment for each stimulus and plotted on the ordinate of the graph. Abscissa values are the average magnitude estimates assigned to each stimulus.

Fig. 4. Stimulus generalization to pure tone intensity following discrimination training. Each extreme stimulus was correlated with reinforcement for one of two responses. Only the R_1 gradient is plotted since the two gradients are complementary. (Magnitude estimation preceded discrimination training.) Each point is the arithmetic mean of ten responses by each of ten subjects.

Fig. 5. Stimulus generalization to pure tone intensity following discrimination training. Each extreme stimulus was correlated with reinforcement for one of two responses. Only the R_1 gradient is plotted since the two gradients are complementary. (Conditioning and generalization testing preceded magnitude estimation.) Each point is the arithmetic mean of ten responses by each of ten subjects.

Fig. 6. Magnitude estimations of pure tone intensities for the group that subsequently received conditioning and generalization testing. Each point represents the geometric mean of 10 estimates from each of 10 subjects, expressed in log relative units.

Fig. 7. Magnitude estimations of pure tone intensities for the group in which conditioning and generalization testing were received previously. Each point represents the geometric mean of 10 estimates from each of 10 subjects, expressed in log relative units.

Fig. 8. Relationship between magnitude estimations and category estimations, constructed from the two-response generalization gradients for pure tone intensities, when magnitude estimations precede conditioning and generalization testing.

Fig. 9. Relationship between magnitude estimations and category estimations, constructed from the two-response generalization gradients for pure tone intensities, when conditioning and generalization testing precede magnitude estimation. The relative frequency of R_1 at each stimulus value was multiplied by unity and that of R_2 was multiplied by two. The arithmetic sum of these weighted frequencies is taken as the category assignment for each stimulus and plotted on the ordinate of the graph. Abscissa values are the average magnitude estimates assigned to each stimulus.

Fig. 10. Stimulus generalization to pure tone intensity following discrimination training. The 50, 75, and 100 db stimuli were correlated with reinforcement for an R_1, R_2 or R_3 response, respectively. Magnitude estimations were determined before conditioning and generalization testing. Each point is the arithmetic mean of ten responses by each of ten subjects.

Fig. 11. Stimulus generalization to pure tone intensity following discrimination training. The 50, 75, and 100 db stimuli were correlated with

reinforcement for an R_1, R_2 or R_3 response, respectively. Conditioning and generalization testing preceded magnitude estimation. Each point is the arithmetic mean of ten responses by each of ten subjects.

Fig. 12. Magnitude estimations of pure tone intensities for the group that subsequently received conditioning and generalization testing. Each point represents the geometric mean of 10 estimates from each of 10 subjects, expressed in log relative units.

Fig. 13. Magnitude estimations of pure tone intensities for the group that received conditioning and generalization testing previously. Each point represents the geometric mean of 10 estimates from each of ten subjects, expressed in log relative units.

Fig. 14. Relationship between magnitude estimation and category estimation, constructed from the three-response generalization gradients for pure tone intensities. Magnitude estimation preceded conditioning and generalization testing. The relative frequency of R_1 at each stimulus value was multiplied by unity, that of R_2 by two, that of R_3 by three. The sum of these weighted frequencies is the category assignment for each stimulus, plotted on the ordinate of the graph. Abscissa values are the average magnitude estimates assigned to each stimulus.

Fig. 15. Relationship between magnitude estimation and category estimations, constructed from the three-response generalization gradients for pure tone intensities. Conditioning and generalization testing preceded magnitude estimation. The relative frequency of R_1 at each stimulus value was multiplied by unity, that of R_2 by two, and that of R_3 by three. The sum of these weighted frequencies is the category assignment for each stimulus, plotted on the ordinate of the graph. Abscissa values are the average magnitude estimates assigned to each stimulus.

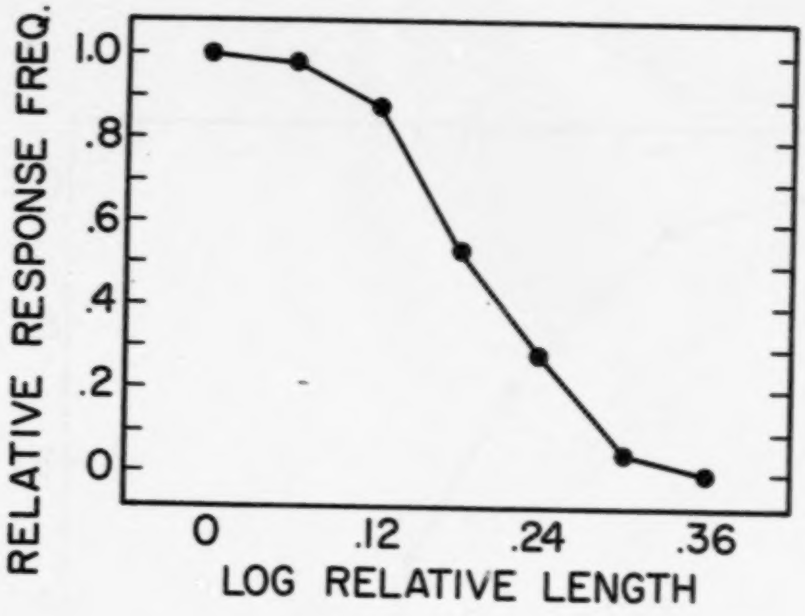

RELATIVE RESPONSE FREQ.
1.0
.8
.6
.4
.2
0
0
.12
.24
.36
LOG RELATIVE LENGTH

Fig. 1

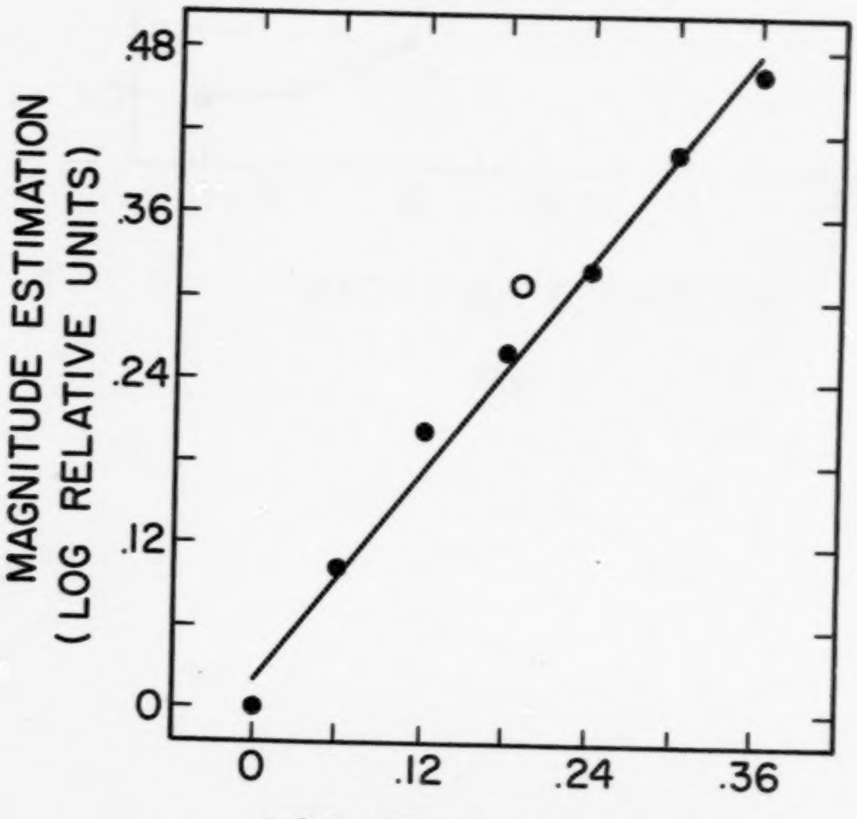

MAGNITUDE ESTIMATION
(LOG RELATIVE UNITS)
.48
.36
.24
.12
0
0
.12
.24
.36
LOG RELATIVE LENGTH

Fig. 2

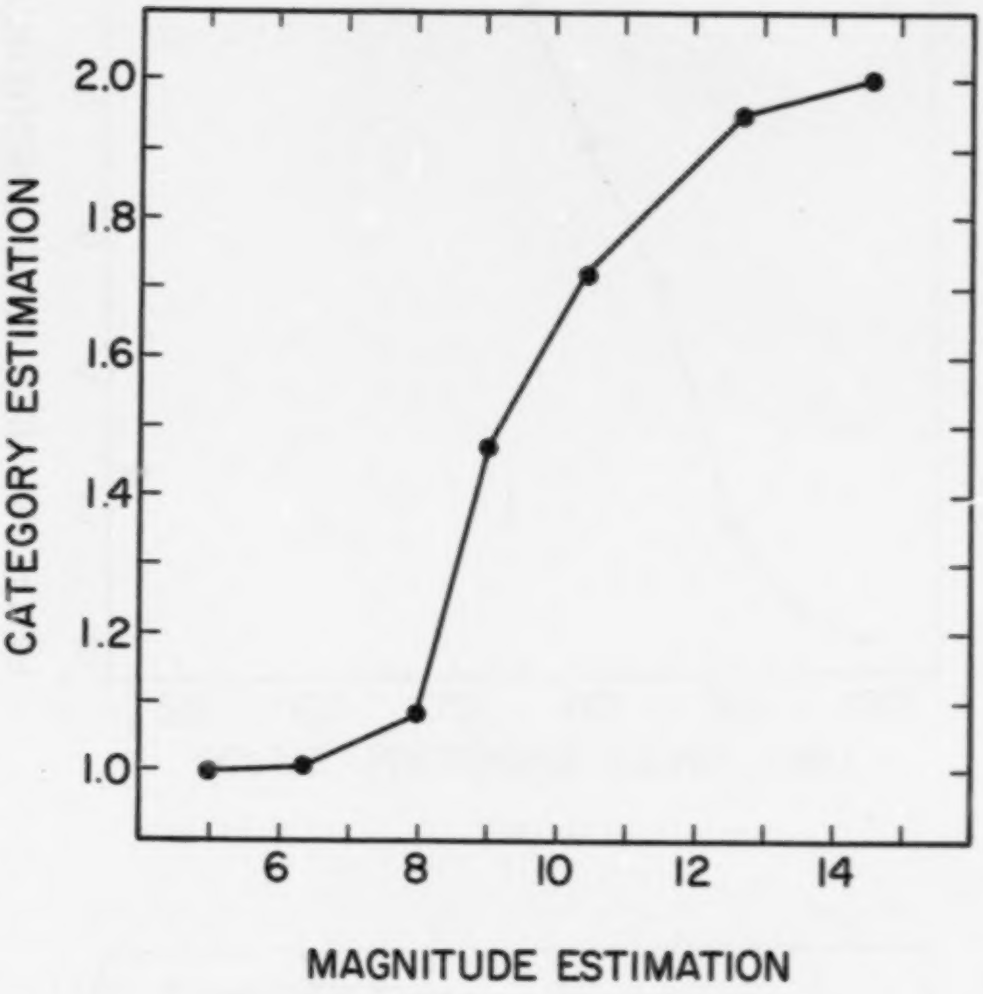

CATEGORY ESTIMATION
2.0
1.8
1.6
1.4
1.2
1.0
6
8
10
12
14
MAGNITUDE ESTIMATION

Fig. 3

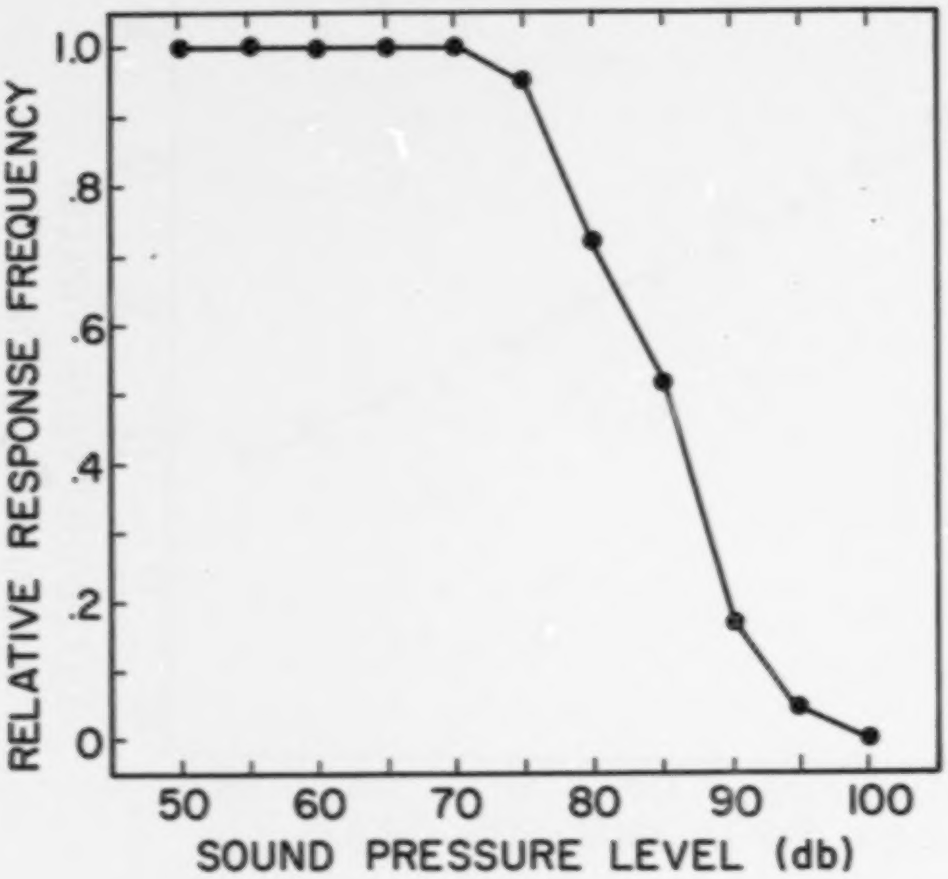

RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
50
60
70
80
90
100
SOUND PRESSURE LEVEL (db)

Fig. 4

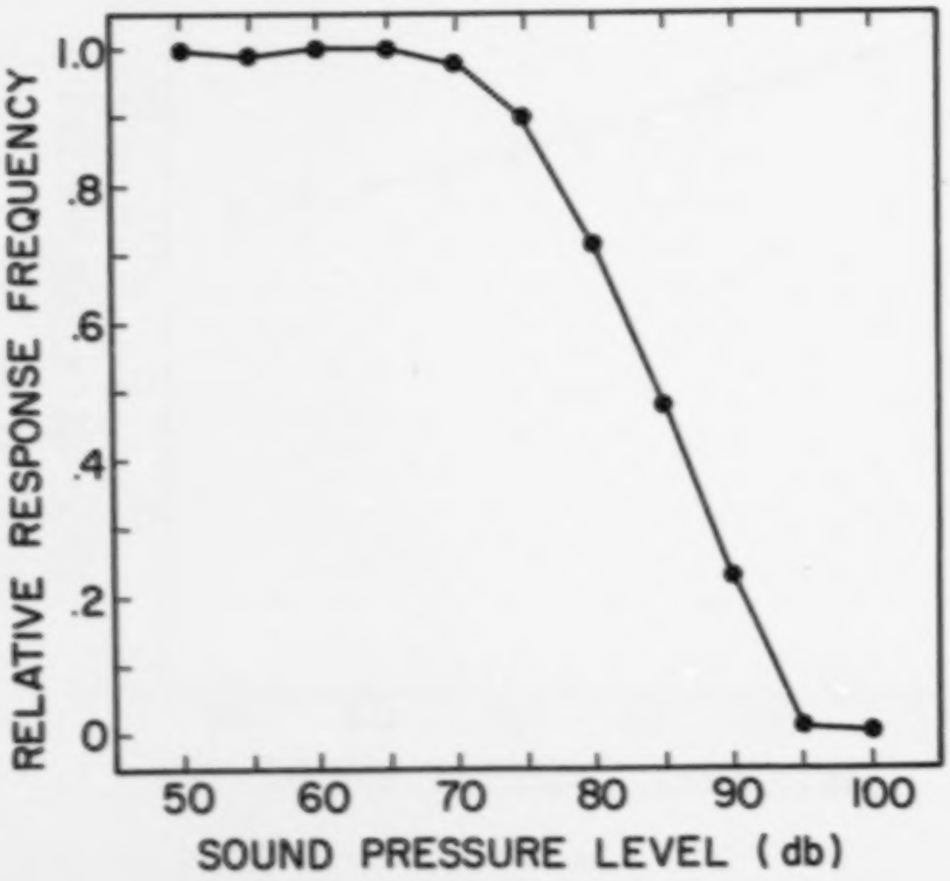

RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
50
60
70
80
90
100
SOUND PRESSURE LEVEL (db)

Fig. 5

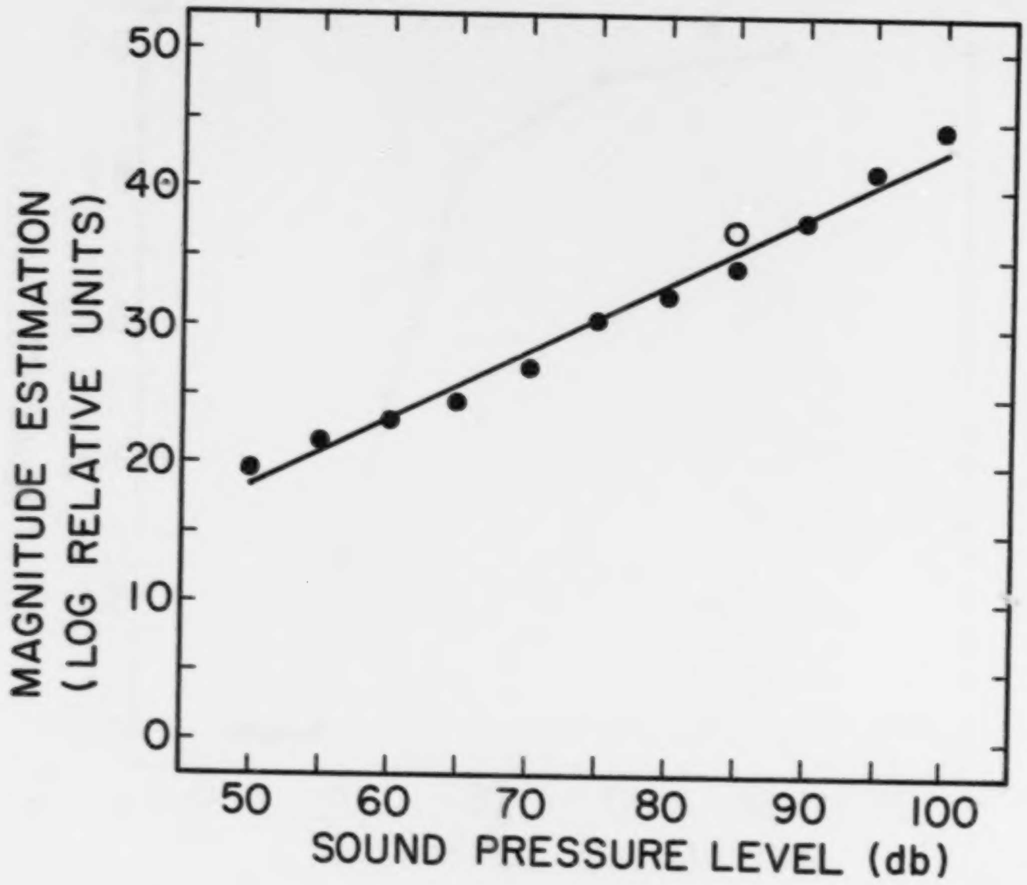

MAGNITUDE ESTIMATION (LOG RELATIVE UNITS)
0
10
20
30
40
50
SOUND PRESSURE LEVEL (db)
50
60
70
80
90
100

Fig. 6

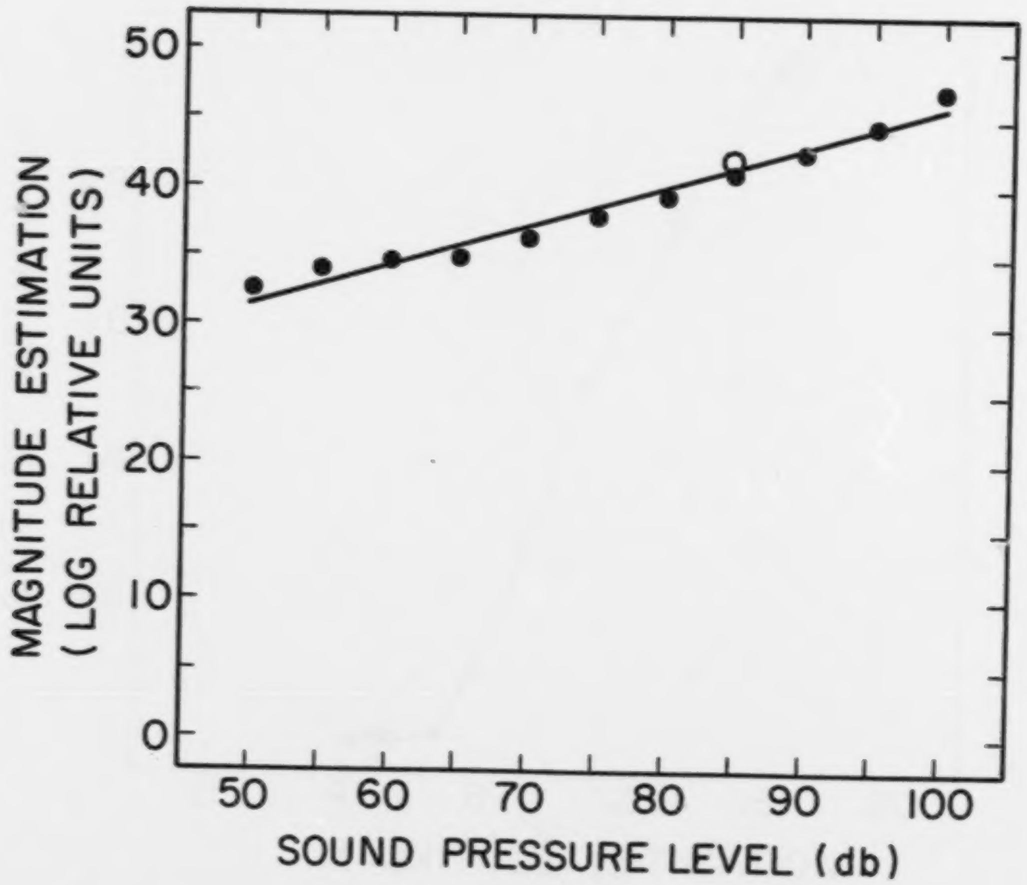
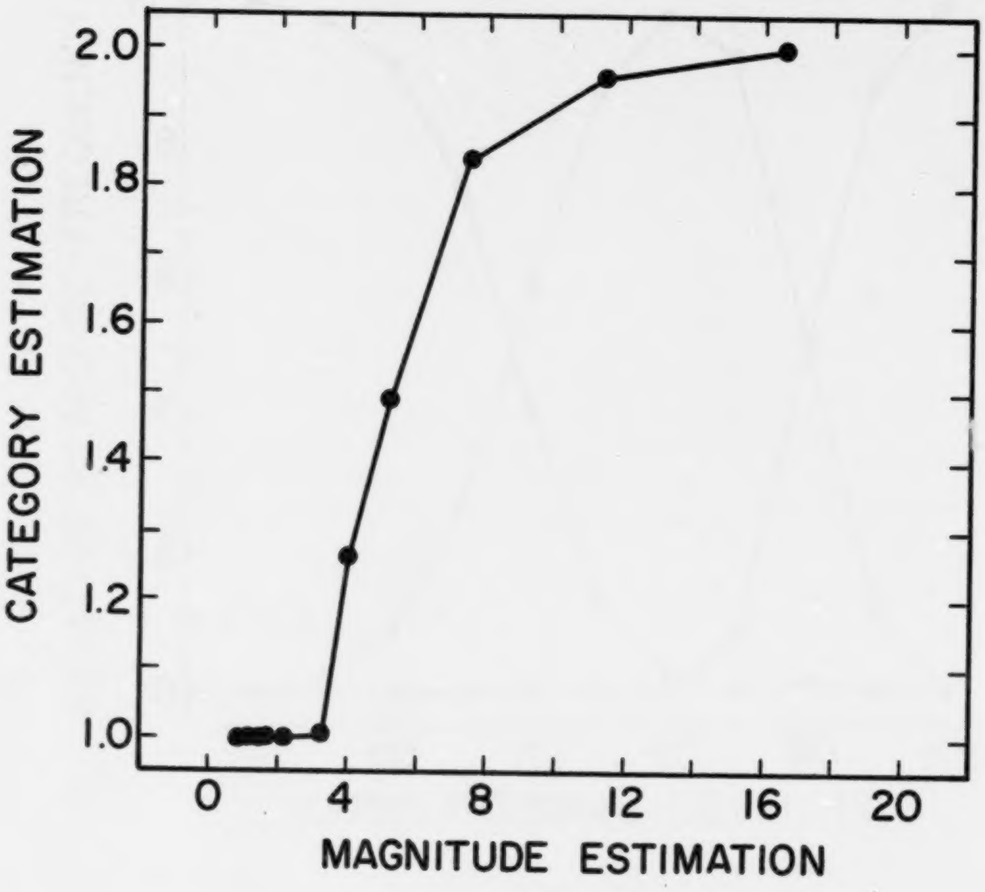
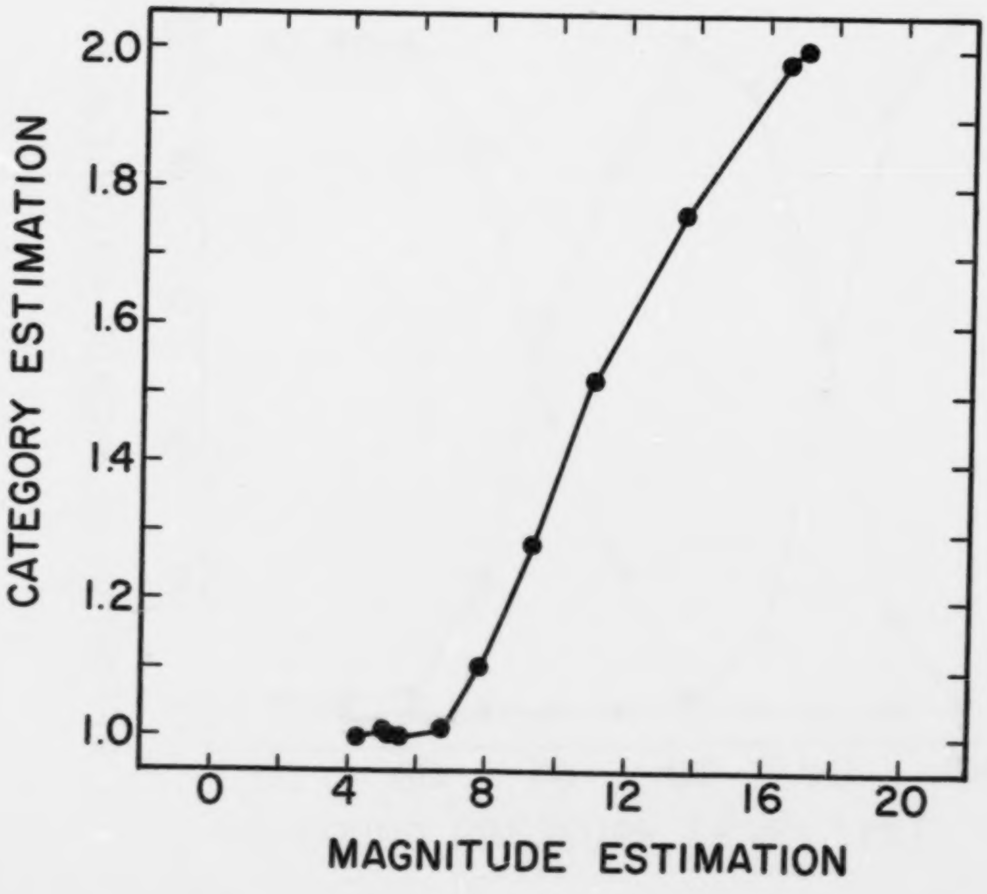

MAGNITUDE ESTIMATION (LOG RELATIVE UNITS)
0
10
20
30
40
50
SOUND PRESSURE LEVEL (db)
50
60
70
80
90
100

Fig. 7



2.0
1.8
1.6
1.4
1.2
1.0
CATEGORY ESTIMATION
0
4
8
12
16
20
MAGNITUDE ESTIMATION

Fig. 8



2.0
1.8
1.6
1.4
1.2
1.0
CATEGORY ESTIMATION
0
4
8
12
16
20
MAGNITUDE ESTIMATION

Fig. 9

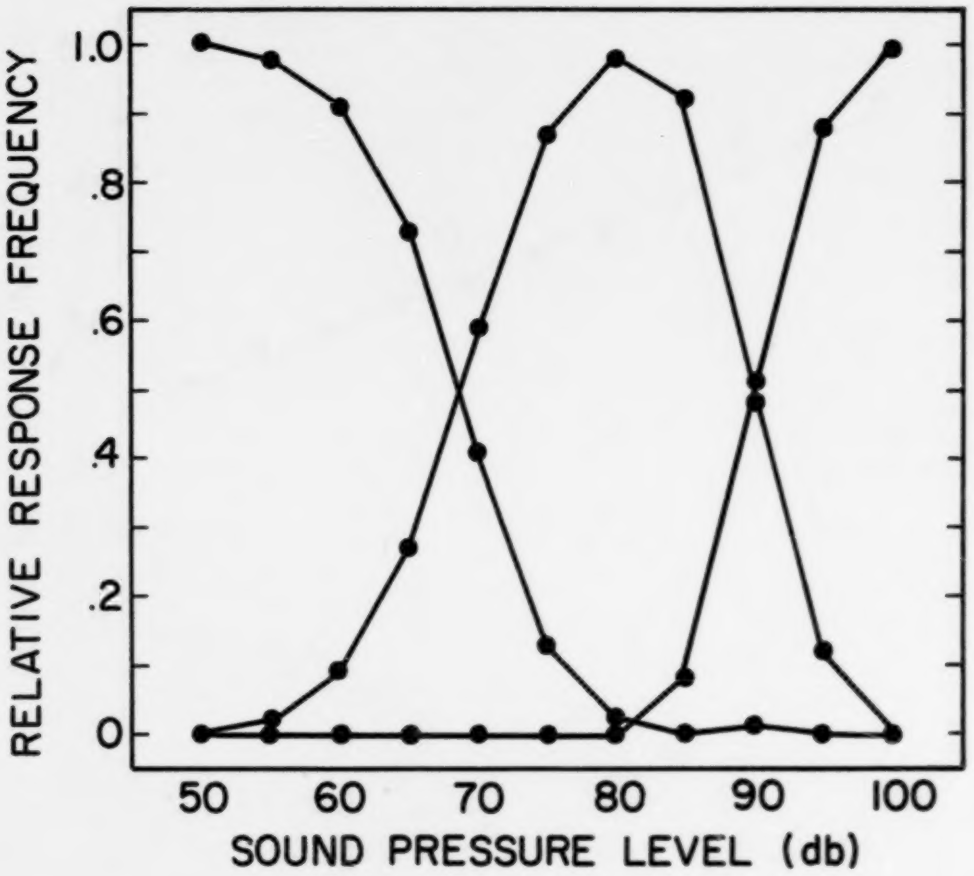

RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
50
60
70
80
90
100
SOUND PRESSURE LEVEL (db)

Fig. 10

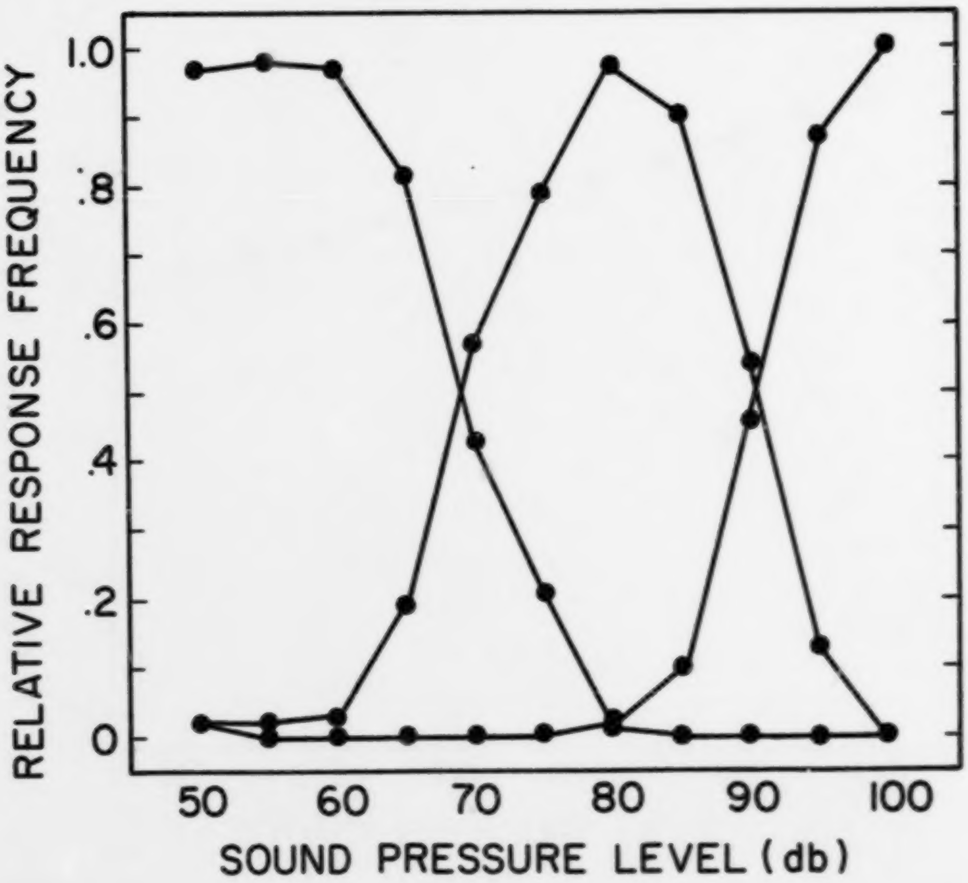

RELATIVE RESPONSE FREQUENCY
1.0
.8
.6
.4
.2
0
50
60
70
80
90
100
SOUND PRESSURE LEVEL (db)

Fig. 11

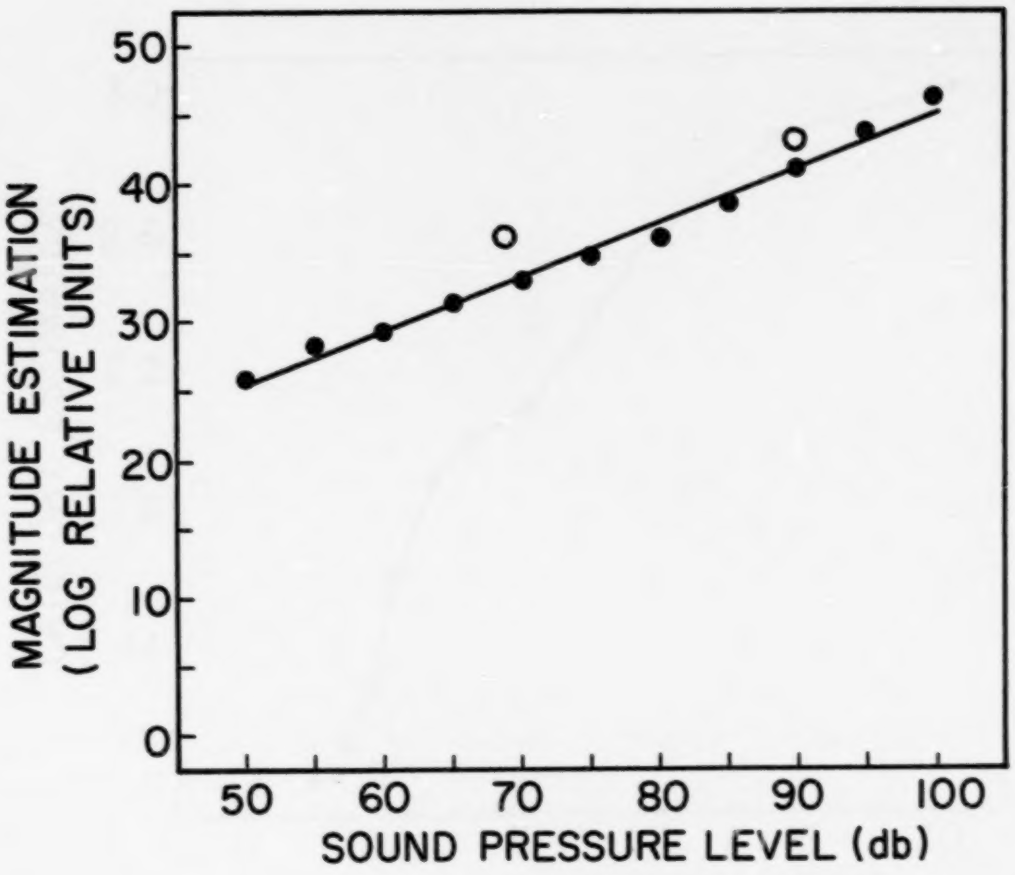

MAGNITUDE ESTIMATION
(LOG RELATIVE UNITS)
50
40
30
20
10
0
50
60
70
80
90
100
SOUND PRESSURE LEVEL (db)

Fig. 12

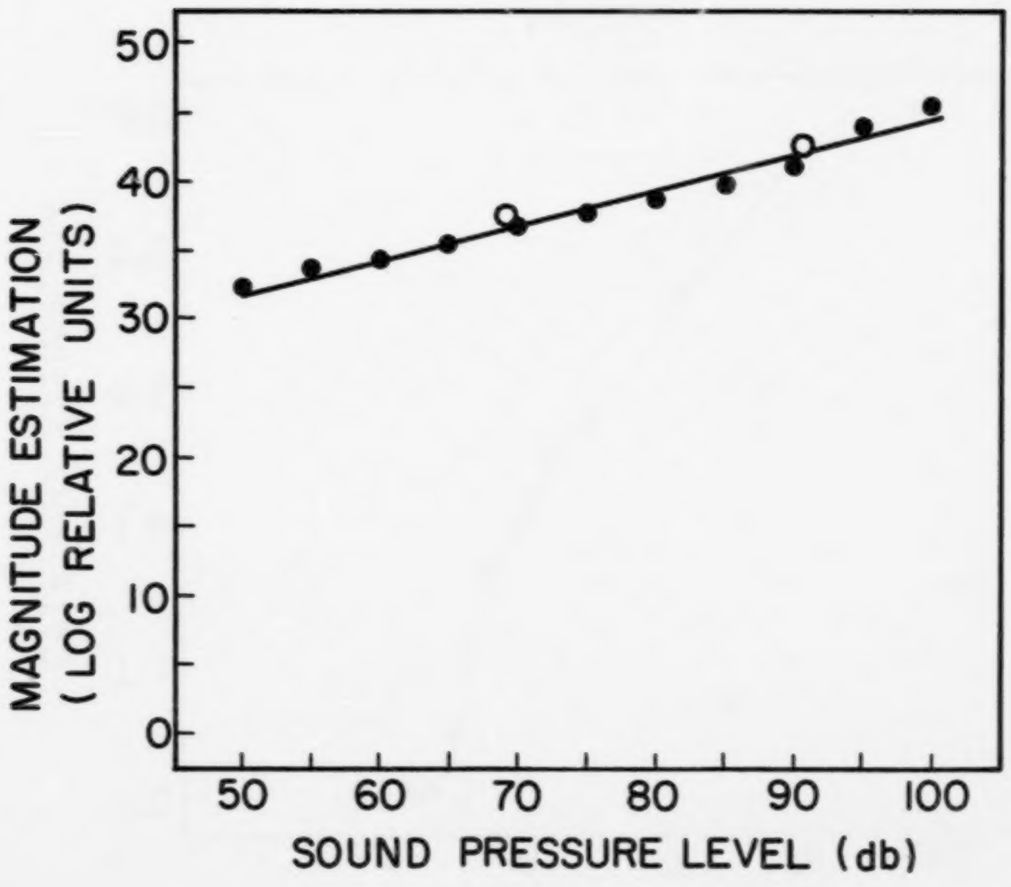

MAGNITUDE ESTIMATION
(LOG RELATIVE UNITS)
50
40
30
20
10
0
50
60
70
80
90
100
SOUND PRESSURE LEVEL (db)

Fig. 13

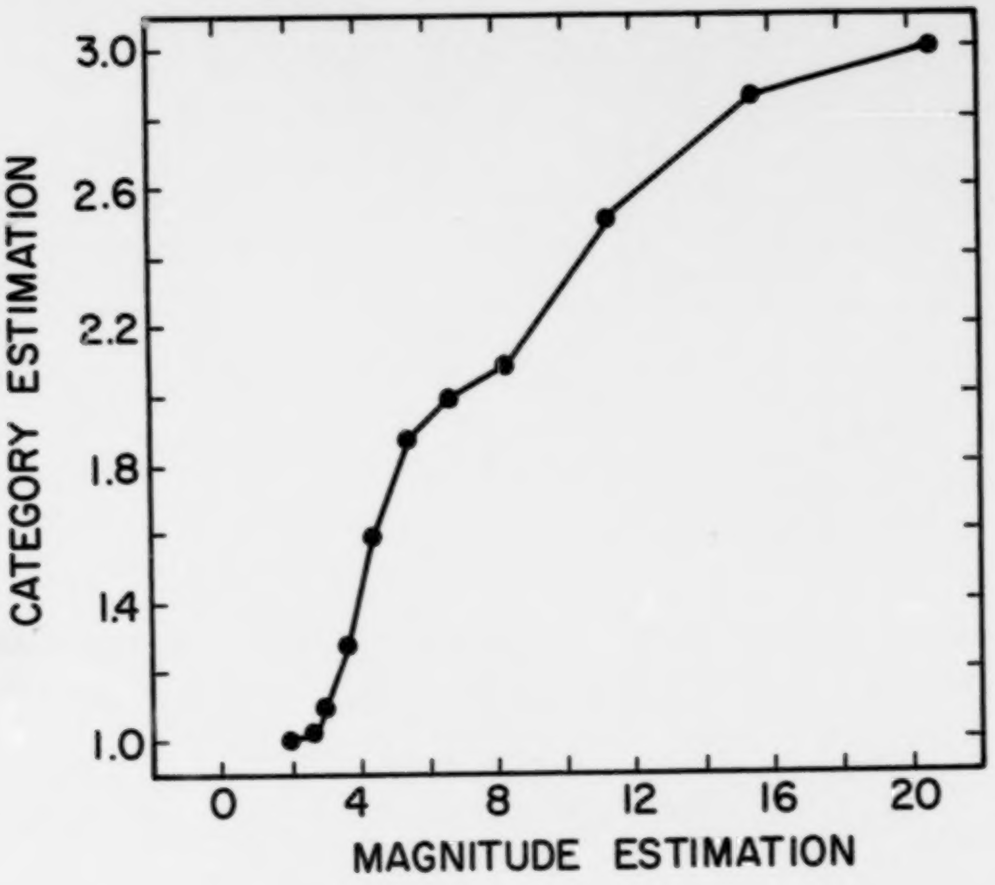

CATEGORY ESTIMATION
3.0
2.6
2.2
1.8
1.4
1.0
0
4
8
12
16
20
MAGNITUDE ESTIMATION

Fig. 14

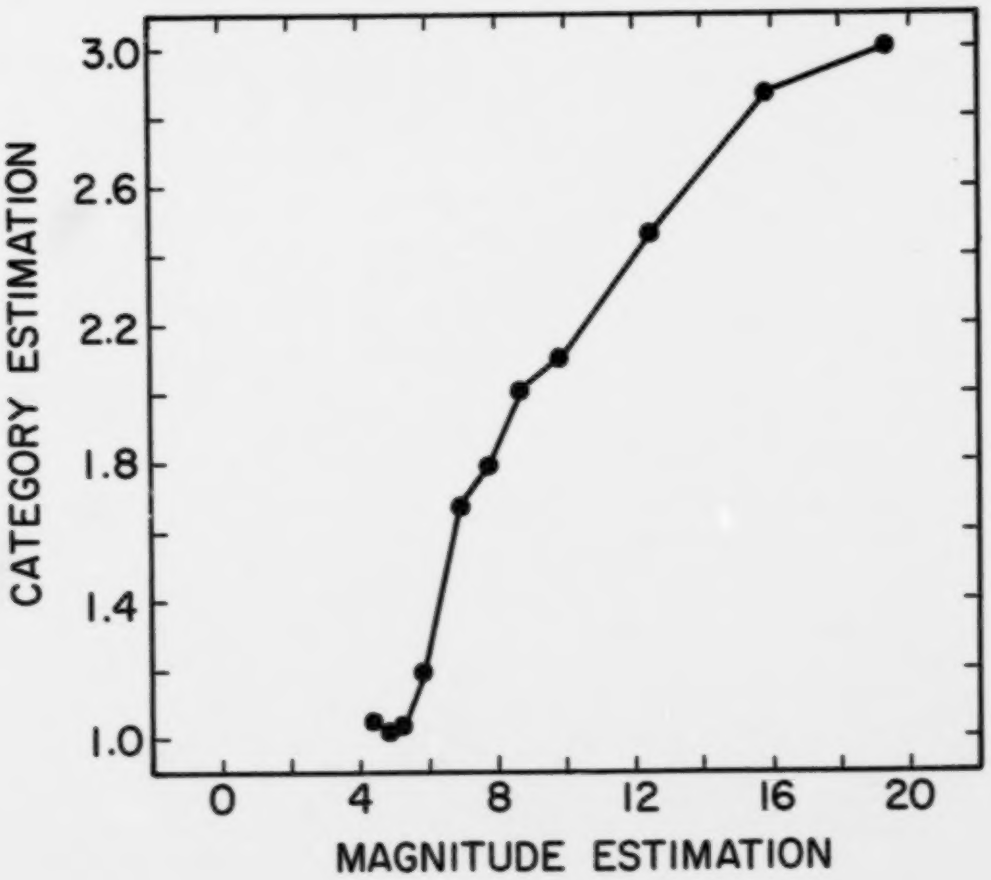

CATEGORY ESTIMATION
3.0
2.6
2.2
1.8
1.4
1.0
0
4
8
12
16
20
MAGNITUDE ESTIMATION

Fig. 15

AN APPLICATION OF MEAN VALUE THEORY TO PSYCHOPHYSICAL MEASUREMENT

D. V. Cross
The University of Michigan

An application of mean value theory to psychophysical measurement

D. V. Cross

Theory of measurement

Pfanzagl's theory of measurement.

- Representation theorem.
- Uniqueness theorem.

Application to psychophysics: bisection experiments

Empirical test of bisymmetry.

- Apparatus.
- Procedure.
- Results.

The construction of individual bisection scales.

- Apparatus and procedure.
- Results.
- The scaling procedure.

Generalization to operations involving more than two variables

The generalized mean.

- Formal properties.
- Applications.

Effects of prior stimulation on mean value levels.

Characterization of mean values for arbitrary functions.

- Representation theorem for nonsymmetric means.
- Four properties of mean values for arbitrary functions.

An application of mean value theory
to psychophysical measurement

D. V. Cross

Behavior Analysis Laboratory

Theory of measurement

The physicist N. P. Campbell defined measurement, in its widest sense, as "the assignment of numerals to things so as to represent facts or conventions about them" (Ferguson *et al.*, 1939-40, p. 340). This was a natural definition since the physical sciences deal primarily with additive magnitudes such as "mass" and "length" and the measurement of objects with respect to these properties is well founded on the availability of additive empirical operations. A mapping of a set S (of physical objects) onto the set of real numbers can be performed in such a way that inferences about relations between elements of S can be drawn from corresponding relations between their assigned numbers. In particular, the *fact* that the operation of combining two elements *a* and *b* results in an effect on a physical system equal to that of a single third element, *c*, is easily represented by the convention of assigning a number to *c* that equals the sum of the numbers assigned to *a* and *b*. A rigorously axiomatized theory of measurement for additive magnitudes was developed as early as 1901 by Hölder. The axioms specify the precise properties an empirical operation must have in order to be represented by the arithmetical operation of addition in the real number system.

Since additive empirical operations are virtually nonexistent in psychology, it has been argued that "genuine" measurement (i.e., procedures leading to ratio or interval scales of psychological magnitudes) is impossible and that our procedures lead only to something on the level of an ordinal scale. For years this criticism was supported by the lack of an adequate theory of measurement as well founded as the theory for measuring additive magnitudes. The current situation, however, is quite different. Rigorous foundations for psychological measurement have been incontestably laid. As pointed out by Suppes and Zinnes (in Luce, Bush and Galanter, 1963, Chapter 1) the fundamental problem in psychological measurement is not the absence of additive empirical operations but rather the problem of formally characterizing the empirical operations employed in a given experiment and the type of relations obtained, and showing that these are equivalent to the operations and relations of an appropriately chosen numerical system. In this sense, fundamental measurement procedures have been established that are not based on an additive operation but that nonetheless lead to interval and even ratio scales (e.g. Pfanzagl, 1959; Luce & Tukey, 1964).

Pfanzagl's theory of measurement.

Pfanzagl (1959) presented an especially interesting axiom system for measurement. It will be discussed in some detail here because it has application to a variety of measurement problems in psychology—from the construction of scales of sensory function based on bisection procedures to the scaling of subjective utility according to the concept of von Neumann and Morgenstern. A specialization of the theory leads to the classical model for measuring additive magnitudes mentioned above. Actually the theory is a reinterpretation of Aczél's (1948) axiomatization of mean values about which more will be said later in application to adaptation-level theory.

Representation theorem. If a binary operation "o" defined on a weakly ordered and connected set M satisfies the following conditions:

(i) existence: for each pair of elements a, b $\in$ M there exists a unique element aob $\in$ M

(ii) monotonicity: if a $\overset{<}{\underset{>}{\sim}}$ a' then aob $\overset{<}{\underset{>}{\sim}}$ a'ob and boa $\overset{<}{\underset{>}{\sim}}$ boa' for all b $\in$ M.

(iii) continuity: the operation aob is continuous in both a and b.

(iv) bisymmetry: $(aob)o(cod) \sim (aoc)o(bod)$.

then and only then there exists three real numbers p, q, and r $(p, q > 0)$ and a continuous and strictly monotonic function $f{:}M \rightarrow R$ which maps the set M into the set R of real numbers such that

$$f(aob) = pf(a) + qf(b) + r. \tag{1}$$

Uniqueness theorem. The mapping f is unique up to a positive linear transformation, that is, if (1) holds for g as well as f, then

$$g(x) = af(x) + b \qquad (a,b \text{ constants}, a > 0); \tag{2}$$

thus the numerical assignment given by f is an interval scale.

Comparison of this theory with axiomatic treatments of the theory of additive measurement (e.g., Ross, 1964, p. 51) reveals some obvious similarities. Both postulate a weak order on the set M, that is, a binary relation defined on the elements of M that is

i. transitive: if $a \lesssim b$ and $b \lesssim c$ then $a \lesssim c$

ii. connected: either $a \lesssim b$ or $b \lesssim a$ or both, and

iii. antisymmetric: if $a \lesssim b$ and $b \lesssim a$ then $a \lesssim b$.

Empirical procedures and models leading to a weak ordering of elements of M are discussed in detail by Coombs (1964). Both theories also postulate a binary operation that maps MXM onto M monotonically and continuously in both of its arguments. However, the theory of additive measures assumes that the operation is "associative", i.e, $(aob)oc = ao(boc)$, and "commutative", i.e., $aob = boa$. The present theory drops the commutativity requirement altogether and instead of associativity it asserts that the function of any four variables

$$\psi(x_{11}, x_{12}, x_{21}, x_{22}) = F[F(x_{11}, x_{12}), F(x_{21}, x_{22})]$$

does not alter if we replace x_{12} by x_{21} and vice versa. Most of the mathematical power in the theory is contained in this condition. The scaling problem is reduced to that of finding a mapping $a \to a' \in R$ which preserves the initial ordering of M and which defines the function $F(a',b') = (aob)'$ such that

$$F[F(a',b'),F(c',d')] = F[(a',c'),F(b',d')].$$

Aczél (1948) has proved that the most general strictly monotonic and continuous solutions of this functional equation have the form

$$F(a',b') = \phi^{-1}[p\,\phi(a') + q\,\phi(b') + r].$$

Thus, by defining $f(x) = \phi(x')$ we obtain (1) by $f(aob) = f[F(a',b')] = p\,\phi(a') + q\,\phi(b') + r = pf(a) + q\,f(b) + r$. Since all functions satisfying (1) are linearly related as in (2), the constants or "weights" p and q are uniquely determined by the operation o but the constant r is not. For if $f(xoy) = pf(x) + qf(y) + r$ and $g(xoy) = sg(x) + tg(y) + u$ and $g(z) = af(z) + b$, then $apf(x) + aqf(y) + ar = asf(x) + sb + atf(y) + tb + u - b$; and thus $s=p$, $t=q$, $u = ar + b(1 - s - t) = ar + b(1 - p - q)$. In the general case, where $p + q \neq 1$, there exists a natural origin and the scale is unique up to multiplication by a constant, since we can choose a value for b so that $u = 0$, namely $b = ar/(p + q - 1)$. Hence, for $g(x) = f(x) + r/(p + q - 1)$ and for every $h(x) = ag(x)$

$$(3) \quad g(xoy) = pg(x) + qg(y) \qquad (p + q \neq 1).$$

From this result we can see the relation of Pfanzagl's theory to the classical measurement model, for if xoy satisfies, instead of bisymmetry, the stronger conditions of associativity and commutativity then $p = q = 1$ and, from (3)

$$(4) \quad g(xoy) = g(x) + g(y)$$

Application to psychophysics: bisection experiments

In the method of bisection the subject is typically presented three stimuli a, x, and b. The terminal stimuli a and b are fixed and the middle stimulus x is continuously variable. The subject is asked to adjust x until it is 'half the distance between a and b', or it is 'as much like a as it is like b' or until he cannot decide whether it is 'nearer a or b', and so on, depending on the experiment.

Bisection data have been used to construct scales of subjective pitch (Stevens & Volkman, 1940) and loudness (Stevens, 1936; Garner, 1954) on the assumption that the bisection stimulus divides the sensory continuum into equal intervals. The validity of bisection scales have been questioned. Restle (1961) pointed out that subjects may not set their bisection adjustments at halfway but instead set them at some other fractional distance in which case it would be " . . . impossible to construct a consistent scale from bisection data" (p. 203). In bisection experiments involving pure tones varying in intensity, large differences are obtained in the bisection point adjusted by the subject, depending on the order in which the tones are presented. If they are presented in ascending order from faintest to loudest, the bisection point is typically set higher than that set when the tones are presented in descending order. This order effect is called hysteresis (Stevens, 1957). Hysteresis was also found in experiments on pitch bisection (Cohen, Hansel, and Sylvester, 1954). An empirical test of the internal consistency of the bisection method as applied to scaling loudness was reported by Gage (1934). Five stimuli were involved. The largest interval was bisected first, producing $a_1 o a_5 = a_3$ as the bisection point. Then the intervals (a_1, a_3) and (a_3, a_5) were bisected, yielding the bisection points $a_1 o a_3 = a_2$ and $a_3 o a_5 = a_4$. The subjects were then asked to bisect the interval (a_2, a_4), producing a stimulus $a_2 o a_4 = a_3'$. It was expected—if bisection results are internally consistent—that a_3' should equal a_3. The results were to the contrary. This constitutes a test of bisymmetry if the bisection operation is both commutative and reflexive (i.e., if $aob = boa$ and $aoa = a$) for then and only then

$$a_3' = a_2 o a_4 = (a_1 o a_3) o (a_3 o a_5) = (a_3 o a_1) o (a_3 o a_5) = (a_3 o a_3) o (a_1 o a_5) = a_3 o a_3 = a_3.$$

The large order effects obtained in the bisection of loudness intervals is proof, however, that the operation is not commutative in this application. The large order effects typically obtained in bisection experiments prove that the commutativity assumption is invalid, but bisection may still be a metric operation in the sense indicated by the measurement theorem described above. In fact since bisection operations are probably reflexive, additional restrictions can be placed on the constants p, q, and r in (1), namely $r=o$ and $p+q=1$. This can be shown as follows: let x and y be the least and greatest stimuli, respectively, in the interval to be bisected, and let $f(x) = 0$ and $f(y) = 1$; if, in equation (1), $a = b = x$, then $r = 0$; if $a = b = y$, then $p + q = 1$. In addition, by $a = x$ and $b = y$, $f(xoy) = q$ and by $a = y$ and $b = x$, $f(yox) = p$. It is clear that only if the assumption of commutativity is valid does $p = q = 1/2$.

Thus, if bisections are reflexive as well as unique, strictly monotone, continuous and bisymmetric, then they satisfy the necessary and sufficient conditions for weighted mean values (cf. Aczél, 1948). An interpretation suggested in application of the model is that the subject adjusts the variable stimulus in the a, x, b sequences to a value that evokes a sensory effect (e.g., loudness) equal to a weighted average of the effects evoked by a and b respectively. The "weights" corresponding to the constants p and q in (3), may be interpreted as representing the relative contributions of the first and last stimulus in the sequence toward the determination of the bisection point. The differential weighting of the terminal stimuli in the scaling procedure permits the order effect mentioned above, wherein the bisection point is biased in the direction of the most recent stimulus in the sequence, whether ascending or descending.

Empirical test of bisymmetry.

A direct experimental evaluation of the validity of the bisymmetry assumption for the bisection operation has never been reported. A proper test involves, to begin with, at least four distinct stimulus values a, b, c and d. The bisection points obtained for the (a,b) and (c,d) intervals in turn define a new interval which must be bisected. Under the hypothesis of bisymmetry, the resulting bisection point should equal—allowing for measurement errors—the bisection point of that interval defined by the bisection points of the (a,c) and (b,d) intervals. This test was performed as part of a preliminary investigation of the bisection procedure as a metric operation for scaling sensory continua.

Apparatus. Figure 1 is a schematic of the apparatus involved in the production and presentation of the pure tones used as stimuli in the experiments described below. The stimulus source was two audio oscillators (Hewlett-Packard 200 AB), one providing the top stimulus in the interval to be bisected, and the other providing both the bottom stimulus and the middle variable stimulus. Three separate gating circuits were employed to switch the tones in correct sequence to the subjects' earphones. The components of these circuits were two electro-optical devices (Raysistor CK1114), each containing a light source and a photocell (see lower left of Fig. 1). A keying voltage applied to the light source produced a change in the resistance of the photocell which permitted the applied signal to pass. These circuits provided distortionless switching of the tones with controlled rise-decay time (approx. 20 msec) and an on/off signal ratio exceeding 120 db. The intensities of the terminal stimuli of a bisection interval were fixed by the output level of oscillator No. 2, for the highest intensity tone, and the output level of oscillator No. 1 (minus the insertion loss provided by the attenuator interposed between oscillator No. 1 and gate No. 1), for the lowest intensity tone. The intensity of the bisection stimulus, x, was continuously variable over a range that depended on the output of oscillator No. 1 and the corresponding attenuator setting. The intensity of x was controlled by the subject (S), who adjusted a three-turn linear potentiometer (Helipot Mod C), which was interposed between oscillator No. 1 and gate No. 2. Four precautions were taken to avoid the biasing effects of position cues in the setting of the bisection point: (1) E systematically altered the range over which x could vary; (2) a three-turn potentiometer was used (1080° rotation from extreme counter-clockwise setting to extreme clockwise setting) which was (3) concealed from the direct view of S; (4) E required that S reset the potentiometer to either extreme position each time a new interval was presented.

A "presentation" consisted of a sequence of three tones, each of 300 msec duration, with 200 msec intervals between tones. The tones occurred in either ascending ($\langle a\ x\ b \rangle$ $a < b$) or descending ($\langle b\ x\ a \rangle$, $b > a$) sequence, depending on the schedule. The S initiated each presentation himself by actuating a switch that keyed the control circuitry, allowing a stepping relay to sequentially activate the gates. The S could listen to the sequence as many times as he wished, by repeatedly pressing the lever, until he was satisfied that his final adjustment bisected the interval. He was encouraged to bracket the bisection point, approaching from above and below, and to make successively finer adjustments until the bisection point was found. The tones were presented to S binaurally, through calibrated headphones (TDH-39 w/neoprene cushions). The tone intensities were measured as the voltages impressed on the headphones, and converted to units of sound pressure-level by reference to the calibration characteristics of the earphones.

Procedure. Two sets of stimuli were involved in separate tests of the bisymmetry assumption. The stimuli were 1000 cps tones at 65, 75, 80, and 90 db SPL (as measured at the headphones) in set A; and 60, 70, 85 and 95 db SPL in set B. Two Ss served in the experiment; both were University students with no previous experience in a psychophysical experiment. They were given preliminary practice sessions in bisecting intervals other than those described below. Subject 1 participated in a total of ten one-hour experimental sessions on separate days, performing successive bisections of intervals from set A and

B in both ascending and descending sequence. Subject 2 served in only 2 sessions, each involving intervals from set B, once in ascending order and the other in descending order. The procedure in each session was as follows:

(1) The primary intervals were bisected first: for example, from set A the intervals (65, 75), (80, 90), (65, 80) and (75, 90) were bisected, yielding bisection points B(65, 75), B(80, 90), B(65, 80) and B(75, 90), respectively. The intervals were presented according to a random schedule, and two determinations were obtained at different times in the schedule for each bisection point. The average (in physical units) of the two determinations for a given interval defined a new stimulus which served as a terminal stimulus for a secondary interval.

(2) The secondary intervals [B(65, 75), B(80, 90)] and [B(65, 80), B(75, 90)] were then bisected. Three determinations for each interval were obtained in random order.

The same steps were followed for the tones presented in reverse order (i.e., in descending sequence) and for the intervals defined by the stimulus intensities in set B. Subject No. 1 bisected ascending sequences of set A intervals in the first four experimental sessions and descending sequences in the second four sessions. He then served in two sessions bisecting intervals from set B in ascending order only. Subject No. 2 bisected intervals from set B only, in ascending order during the first session and in descending order during the second.

Results. The bisections obtained for each interval presented to S 1 in sessions 1-8 are shown in Table 1. In every session except the first, the expected correspondence between the bisections of bisection point intervals appears to hold. Within a given session, S was fairly consistent in his bisections. The largest absolute difference between the two determinations obtained for any interval was 2.0 db and the smallest was 0.2 db with an average absolute difference of 1.1 db for the ascending sequences and 0.6 db for the descending sequences. This can be compared with an absolute difference between the two secondary bisection points for each session (including session I) of 0.5 db. The two secondary bisection points obtained from S 1 with set B stimuli were 86.7 and 86.6 in one session and 86.7 and 86.4 in the other. Although S 2's bisection settings differed noticeably from those of S 1, his secondary bisection points were also close: 82.4 and 82.5 for the ascending sequence and 78.0 and 78.9 for the descending sequence.

A small but consistent hysteresis effect was observed in these results (see Table 1). Bisection settings were higher for ascending sequences than they were for descending sequences: For S 1 the average effect was 1.9 db for 10 db intervals and 2.3 db for 15 db intervals. Since bisymmetry does not appear to be violated, however, bisection should be a metric operation in the sense specified above.

The construction of individual bisection scales for loudness. Previous attempts at the construction of a scale for loudness based on bisection data (e.g., Garner, 1954) employed, not the bisection operation, but rather an equisection procedure, in which S sat before a set of five keys which he pressed to produce the tones. The intensities presented when he pressed the end keys were fixed, and S had to adjust the intensities produced by pressing the intermediate keys in order to divide the given interval into four "equal-appearing" steps in loudness. The equisection procedure is generally considered to be a generalization of the bisection method, since in the construction of equisection scales it is assumed that S makes successive bisections of the interval, dividing it into equal "sense distances" which may be taken as a unit of measurement. This is a much more stringent assumption, and thus one much likelier to be violated, than the assumption that a single bisection of an interval occurs at a stimulus point on the underlying continuum that evokes a sensory effect equal to a weighted mean of the effects evoked by the two terminal stimuli. In the following experiment bisection scales of pure tone intensity were constructed for 2 *S*s using simple bisection data.

Apparatus and procedure. The method of obtaining the bisections was the same as that described above. The stimuli were 1000 cps tones as before, with the following intensity levels: 60, 65, 70, 75, 80, 85, and 90 db (SPL) as measured at *S*'s headphones. Bisections were made of all possible intervals formed by pairing one intensity level from the set with another, including each level paired with itself for an empirical test of the reflexivity assumption. Thus, forty-nine separate pairs of stimulus intensities were bisected (considering ascending and descending sequences of the same two intensities as distinct pairs). Data were obtained from two *S*s: one was *S* 1 in the preceding experiment; the other, *S* 3, was a research assistant and a practiced psychophysical observer. Both *S*s were given extensive practice before the experiment began.

Subject No. 1 served in four one-hour experimental sessions, yielding two determinations of each bisection point. Both ascending and descending bisections were made in each session. Two complete sessions were necessary to obtain a single bisection of each of the 49 possible stimulus pairs. The pairs that were involved, their sequential order (whether ascending or descending) and their order of presentation in the schedule were determined randomly for each session. Subject No. 3 served in only two sessions, yielding a single bisection point for each interval.

Results. The results for both Ss are shown in Table 2. For S 1, the decibel averages of the two determinations obtained for each interval are shown. It is noteworthy that no hysteresis effect was observed in the bisection adjustments of S 3 and the effect present in the adjustments of S 1 is negligible, well within the bounds established by an estimate of his average error in replicating a bisection for a given interval. The average absolute difference between the two bisections of each interval for S 1 was 1.3 db, although for six bisections the error was as great as 3 to 7 db. Decibel averages of the ascending and descending bisections were computed for each interval and the results are shown separately for each S in Figs. 2 and 3. The points in these figures correspond to the cell entries of the symmetric matrices in Table 2. The abscissas represent the intensity of the first tone in the $\langle a\ x\ b \rangle$ sequence, the ordinates represent empirical estimates of the middle or bisection tone, x, and the different curves (solid lines) represent different levels of the third tone, b. The unfilled circles in each figure represent the settings obtained for x when a = b. The Fechnerian assumption that loudness is a logarithmic function of stimulus intensity taken together with the assumption that a bisection point divides the psychological continuum into equal intervals implies that each bisection should occur at the logarithmic mean of the interval (or geometric mean in absolute physical units). Under these assumptions, the points connected by dashed lines in Figs. 2 and 3 should represent equal intensities. In other words, the dashed curves should be parallel and separated (in these graphs) by intervals of 2.5 db. This outcome is obtained for the bisections by S 3 (Fig. 2) for all intervals except those involving a 90 db tone and occasionally those involving an 85 db tone. The results for S 1 are different, however. The dashed curves in Fig. 3 are clearly concave upward, contradicting the prediction that follows from a logarithmic psychophysical law.

The monotonicity condition necessary for the application of the measurement model discussed above is clearly satisfied by the data from S 3. There is only one violation of the condition in the data of S 1, involving the 65-85 db interval; the two bisections obtained for this interval differed by 6 db in ascending sequence and by 3.8 db in descending sequence. Therefore, for the purpose of scaling these results, we assumed that this bisection was overestimated and we arbitrarily corrected the order of the pair of intervals contributing to the violation. It also appears that the bisection operation is reflexive, at least for intensities greater than 65 db. Bisection of the (60, 60) pair was overadjusted by both S 1 and S 3 by 1 db and bisection of the (65, 65) pair was also overadjusted by 1 db by S 1. It is difficult to evaluate whether these discrepancies constitute violations of reflexivity or whether they should be considered within tolerable limits of measurement error.

The scaling procedure. Under the assumptions that the bisection operation is unique, monotone, continuous, bisymmetric, reflexive, and (at least for the data at hand) symmetric, there exists an order preserving function

$$f(a \circ b) = \tfrac{1}{2}[f(a) + f(b)] \tag{5}$$

which constitutes a numerical assignment for the stimuli involved that is unique up to a positive linear transformation. The order defined by the empirical estimates of the bisection intensities for each stimulus pair was taken as a starting point in the analysis for S 1 (with the one correction mentioned earlier). For S 3, smooth curves were constructed through the obtained data points and the order obtained on the points of these curves was used. (This procedure was used because some of the obtained differences between points were as small as 1/10 decibel which could lead to an inconsistent ordering). The problem at hand is to find a numerical assignment for these 49 data points (which correspond to cells of a symmetric 7 x 7 matrix, although in general the procedure applies to asymmetric matrices as well) which preserves their given order and which can be factored into a set of seven measures, one for each row (column) of the matrix such that the number assigned to a given data point (cell) equals one half the sum of the measures assigned to the corresponding row and column. An algorithm for finding an additive representation (when one exists) for finite data matrices was presented by Goode (1964). The procedure is essentially as follows. Since, by the monotonicity assumption, the rows (columns) of the matrix are ordered then the stimuli labelling them (a, b, c,...) must be assigned numbers that are ordered in the same way. If $b > a$, then corresponding to a there is an arbitrary positive number e_0 such that $f(a) = e_0$; since $f(b) > f(a)$, there also exists some arbitrary positive number e_1 such that $f(b) = e_0 + e_1 > f(a) = e_0$. Accordingly, for $c > b > a$ there exists an e_2 such that $f(c) = e_0 + e_1 + e_2 > e_0 + e_1 = f(b)$, and so on recursively. If, by the order given on the cells of the matrix, $f(aoc) > f(bob)$ then (5) implies that $[(e_0) + (e_0 + e_1 + e_2)] = (2\, e_0 + e_1 + e_2) > [(e_0 + e_1) + (e_0 + e_1)] = (2\, e_0 + 2\, e_1)$. By this inequality it is clear that $e_2 > e_1$ and we can introduce a new positive number e_2' such that $e_2 = e_1 + e_2'$. By substitution, $f(c) = e_0 + e_1 + e_2 = e_0 + 2\, e_1 + e_2'$. By proceeding in this manner, the following set of expressions is obtained:

$$\begin{array}{l}
f(x_1) = e_0 + a_{11}\, e_1 + a_{12}\, e_2 + \dots + a_{1n}\, e_n \\
f(x_2) = e_0 + a_{21}\, e_1 + a_{22}\, e_2 + \dots + a_{2n}\, e_n \\
\vdots \\
f(x_n) = e_0 + a_{n1}\, e_1 + a_{n2}\, e_2 + \dots + a_{nn}\, e_n
\end{array}$$

The a_{ij} are obtained by successive substitutions of new quantities e_j in solutions of the sets of inequalities given by the original ordering. Any positive (not necessarily equal) evaluation of the e_j yields a possible solution to this set of equations.

Any solution may be obtained from other solutions by positive linear transformations on the cell measures, and by certain kinds of corresponding transformations on the row (column) measures.

This procedure applied to the ordering assigned to the bisections of S 3 led to the following set of equations:

$$(6)\quad \left\| \begin{array}{c} f(60) \\ f(65) \\ f(70) \\ f(75) \\ f(80) \\ f(85) \\ f(90) \end{array} \right\| = \left\| \begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 3 & 2 & 2 & 1 \\ 1 & 2 & 1 & 6 & 4 & 4 & 2 \\ 1 & 3 & 2 & 10 & 6 & 6 & 3 \\ 1 & 4 & 3 & 14 & 9 & 8 & 4 \\ 1 & 5 & 4 & 19 & 12 & 11 & 5 \\ 1 & 6 & 5 & 24 & 15 & 14 & 7 \end{array} \right\| \cdot \left\| \begin{array}{c} e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{array} \right\| = \left\| \begin{array}{c} 1 \\ 10 \\ 20 \\ 31 \\ 43 \\ 57 \\ 72 \end{array} \right\|$$

The equal-epsilon ($e_j = 1$) solution to these equations shown at the right above, yields numerical assignments for the stimulus levels employed in this experiment. The measures assigned to the separate bisection points by (6) were computed and related to the physical intensities of the obtained bisection points. Figure 4 shows that the assigned measures are approximately linear with the logarithm of stimulus intensity, which agrees with the observation made above that S 3 bisected most of the intervals at the geometric mean of the physical intensities of terminal stimuli. The slight curvature and departure from the logarithmic relation at the top of the graph comes, as observed in Fig. 2, from the bisections of intervals involving 90 db tones.

A more interesting application of the procedure is found in the analysis of S 1's data. The following equations were obtained in solution of the inequalities derived from the ordering of the bisection points:

$$(7)\quad \begin{Vmatrix} f(60) \\ f(65) \\ f(70) \\ f(75) \\ f(80) \\ f(85) \\ f(90) \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 4 & 2 & 2 & 2 \\ 1 & 2 & 3 & 7 & 3 & 4 & 3 \\ 1 & 3 & 4 & 10 & 5 & 6 & 4 \\ 1 & 4 & 6 & 14 & 7 & 9 & 6 \\ 1 & 5 & 7 & 18 & 9 & 11 & 8 \end{Vmatrix} \cdot \begin{Vmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{Vmatrix} = \begin{Vmatrix} 1 \\ 8 \\ 14 \\ 23 \\ 33 \\ 47 \\ 58 \end{Vmatrix}$$

The solution on the right side is again the equal-epsilon ($e_j = 1$) solution. As before, the scale values assigned to the bisection points were computed and related to the physical values corresponding to the obtained bisections. It appears that a function of the form $f(x) = bS^n$ may best express the relation. Figure 5 shows, in log-log coordinates, the relation between the constructed measures and the obtained physical measures for the bisections. Examination of the figure indicates that the function must include an additive constant; this conforms to the lack of an absolute zero point for the constructed scale. Since the obtained scale is unique only up to a positive linear transformation, the parameters in the general expression $\alpha f(x) + \beta = bS^n$ are interdependent. For the present application, the number of parameters can be reduced by dividing through by α :

$$f(x) + \beta/\alpha = \frac{bS^n}{\alpha} \qquad \text{or}$$

$f(x) + \beta' = b'S^n$ and both sides of the equation are expressed in terms of a common unit. An estimate of the additive constant β' was obtained by conventional procedures (Lewis, 1960) and values for the parameters b' and n obtained by the method of least squares. Figure 6 shows the resulting graph, in log-log coordinates, of the equation $f(x) + 25 = 2.457\ S^{0.34}$, where S is sound pressure in microbars. The exponent 0.34 is quite different from the value of 0.6 that Stevens (1955) reports for the sone scale of loudness and it is closer to the value of 0.264 arrived at by Garner (1954), who based his scale in part on equisection data. These results for S 1 were confirmed by the outcome of an experiment using direct estimation procedures. S 1 was asked to estimate the apparent magnitude of 1000 cps tones that varied in intensity from 50 - 100 db (SPL) in 5 db steps. Ten judgments were obtained at each intensity level relative to a 75 db modulus. These data were collected and analyzed by a different experimenter who was not acquainted with any of the previous findings obtained with S 1. The function relating S 1's magnitude estimates to sound pressure level was found to be a power function with an exponent equal to 0.34!

Generalization to operations involving more than two variables

The preceding results are not limited to operations involving only two variables. The assumption that a bisection point evokes a sensory effect equal to a weighted average of the effects evoked by the two stimuli defining the interval bisected can be generalized to other procedures involving more than two stimuli. Consider, for example, an experiment in which S is trained to emit R_1 to S_1, and R_2 to S_2, where S_1 and S_2 are two stimuli sampled from a continuum, and then stimuli intermediate to the two training stimuli are presented in a forced response situation (i.e., S must emit either R_1 or R_2 depending on whether S_i is "nearer" S_1 or S_2). At some point S_i on the continuum the responses will be evoked with equal probability. Where this point occurs is not simply a function of S_1 and S_2 but depends on the range of stimuli presented, the spacing of the S_i, and the relative frequency with which each S_i is presented.

The assumption that the point of equiprobable responding corresponds to the midpoint dividing the (S_1, S_2) interval into two equal intervals often leads to contradictions when the intermediate stimuli are spaced differently, or when they are presented with different relative frequencies. In fact, the relativity of the controlling relations that often obtain between stimuli and responding has posed a serious problem for psychophysical measurement. It has been known for decades that the control exerted by a stimulus over the behavior of an organism (within the context of a single experiment) depends on a host of interacting factors: the general level of stimulation; the range, spacing, and distribution of stimulus values making up the presentation schedule; the order in which the stimuli are presented and temporal parameters connected with their presentation; background sources of stimulation; previous experience of the organism and learning or practice effects occurring during the experiment; the payoff structure of the experiment; and so on. The same stimulus may evoke different, sometimes opposing, responses depending on the context. For example, a tone may be judged "loud" when it is the topmost stimulus in one series and "soft" when it is the bottom stimulus in another series. A grey patch illuminated with monochromatic light may take on the hue of the illuminant when viewed against a dark background but take on the complementary hue when viewed against a light background. One of two equal weights lifted second may seem heavier. All scaling procedures in psychophysics are affected to a greater or less degree by this behavioral relativity. Procedures employed in the measurementof sensory limens are contaminated by series, order, and 'anchor' effects, as in the decentered position of the 'point of subjective equality,' time-order errors and the effects of standards and interpolated stimuli. Procedures leading to psychophysical scales, such as methods of magnitude estimation,and comparative or category rating procedures, are all in varying degrees affected by the total range and general level of stimulation, by the spacing and distribution of the stimuli on the continuum, and by the choice of a standard or modulus stimulus.

The traditional approach in psychophysics is to treat these effects as 'biases and pitfalls' that somehow should be avoided or neutralized by appropriate iterative and counterbalancing techniques. The organism is viewed as a meter or 'transducer of energy' with certain dynamic operating characteristics. It is often held that contextual and order effects simply contribute error to the measurement of these operating characteristics, and that they are best eliminated in practice and theory.

A contrasting view is offered by adaptation-level theorists who contend that these parameters of psychophysical judgments must be taken into account and incorporated in psychophysical laws. Helson (1964) has described this point of view in these words:

> "Behavioral relativity does not imply that the relations between stimulus and response variables are capricious or fortuitous. Rather it makes possible translation of data obtained under one complex of conditions into the frame of reference demanded by other complexes of conditions...Behavioral relativity...is not to be equated with experimental error or with deviations from expected outcomes due to extraneous factors; it should be considered rather as relativity that is lawful, ordered, and predictable from situational and personal factors which influence prevailing adaptation levels" (p. 31).

According to AL theory, upon repeated presentation of a given set of stimuli, the organism becomes 'attuned' or 'adapted' to a certain level of stimulation that represents the weighted mean of three classes of stimuli: (1) stimuli in the immediate focus of attention; (2) background or contextual stimuli; and (3) all stimuli experienced in the past. All responses are referable to this level called the adaptation level because it represents the adaptation or adjustment of the organism to given conditions of stimulation. Stimulus levels below AL evoke one kind of response; stimulus levels above evoke a different, sometimes opposite kind of response, and stimuli at or near AL evoke indifferent or neutral responses. Computation of AL from known stimulus quantities depends upon what function of S, f(S), is assumed or proven to hold. AL theory assumes that the Fechnerian logarithmic function is the prevailing psychophysical law relating psychological magnitudes to physical magnitudes, and AL is defined as a weighted log mean of all relevant stimulation; focal, background and residual. The weighted log mean formula has provided a quantitative method of dealing with many of the problems mentioned above in that the weighting constants employed in the formula represent the relative contributions to level made by various factors affecting the response. However, any weighted mean formulation has this advantage. Thus, a major shortcoming of the quantitative formulation of AL theory is its dependence on the logarithmic relation and the consequent _ad hoc_ solution of empirical constants that are dependent on this assumption. If AL were computed as the arithmetic or the harmonic mean instead of the geometric mean a solution would still be obtained but it would require a different assignment of weights to the various factors under investigation. In general these alternative formulations of AL theory are subsumed by the concept of the generalized weighted mean value, which is characterized by a single parameter, and applies to logarithmic, power or linear forms of the psychophysical function.

The generalized mean.

Formal properties. The weighted mean of order t for the n positively valued variables $x_1, \ldots, x_n$ is defined as

(8) $$M_t(x_1,\ldots,x_n) = \left[\frac{\sum_{i=1}^{n} p_i x_i^t}{\sum_{i=1}^{n} p_i}\right]^{1/t}$$

$M_t(x)$ is a continuous function of t for $-\infty \leq t \leq +\infty$. For t=1 we get the simple weighted arithmetic mean. For t=-1 we get the harmonic mean and for t=2 we get the "root-mean-square." Other special means arise as t approaches certain limiting cases:

(9) $$\lim_{t\to\infty} M_t(x) = \max_i \left\{x_i\right\}$$

(10) $$\lim_{t\to-\infty} M_t(x) = \min_i \left\{x_i\right\} \quad \text{and}$$

(11) $$\lim_{t\to 0} M_t(x) = \left[\prod_{i=1}^{n} x_i^{p_i}\right]^{1/\Sigma p_i}$$

which is the geometric mean. Several interesting properties of the generalized mean are given in Hardy, Littlewood, and Polya (1959). Of special interest is the following property: if $t < s$ then $M_t(x) < M_s(x)$ unless the x_i are all equal, or $s \leq 0$ and an x_i is zero. From this there arises the inequality $\min_i \left\{x_i\right\} < M_t(x) < \max_i \left\{x_i\right\}$ for all finite t, unless the x_i are all equal, or $t \leq 0$ and an x_i is zero. Thus a mean value is bounded from below by the least element in the set and from above by the greatest. Suppose we know that the function relating psychological magnitude to physical magnitude is either a power function (with exponent t) or a logarithmic function. Consider an arbitrary sampling of stimuli from an interval of a continuum bounded by S_1 and S_2. If these stimuli are presented in a two-alternative forced-response identification task, for which the identification function assigns R_1 to S_1 and R_2 to S_2, then the point on the continuum where the two response functions (relative frequency) cross over, (the point of equal response probability) corresponds to the mean value of the psychological magnitudes associated with the different levels of stimulation. If this cross-over point lies to the right of the geometric mean of all stimulus values, then $t > o$ and a power function with a positive exponent should describe the underlying psychophysical relation. The greater the discrepancy (relative to the size of the interval and the spacing of stimuli within it) between the geometric mean and obtained mean, the greater the value of the exponent. If the cross-over point occurs to the left of the geometric mean, then a power function with a negative exponent is indicated. Of course, if it corresponds to the geometric mean, then a logarithmic relation obtains. Stevens (1955) reported the results of a bisection experiment in which Ss were required to adjust the intensity of one tone to the midpoint of the loudness interval between two other tones. The obtained settings were compared to those predicted on the assumption that a power function with an exponent equal to 0.6 relates the loudness of pure tones to their sound pressure levels. The obtained bisections were, in every case, greater than the geometric mean but less than that predicted. This outcome may be interpreted, in view of the formal analysis presented above, as indicating that the power function that describes the bisection of loudness has a smaller exponent than the power function that describes the fractionation and direct estimation of loudness.

Applications. In several experiments performed in this laboratory, two distinct responses were first conditioned to two discriminative stimuli bounding some interval on a continuum, and then intermediate stimuli were presented repeatedly and the relative frequencies of the responses were recorded. Various physical continua have been employed in these experiments on stimulus generalization: pure tone intensity and duration, noise intensity, length of lines, and intensity of electric shock. In each case, the point on the continuum corresponding to equal response probability was compared to the psychological midpoint of the interval between the training stimuli, which was predicted from the power function relating apparent magnitude to physical magnitude for the continuum involved. The values for the exponents of the various power functions used in these predictions were those reported by Stevens (1960). In every case the obtained point fell between that predicted from the relevant psychophysical function and the geometric mean of the values of the two training stimuli. The finding that the midpoint of the psychological interval defined by the two training stimuli was not a good predictor of the point of intersection of the generalization functions led us to consider whether a weighted mean value of all stimuli presented during the experiment, described in psychological units, would be a more successful predictor. In another experiment relating parameters of stimulus generalization to magnitude estimations of pure tone intensity (Kopp & Cross, 1964) Ss were trained to press one button if a 1000 cps tone was presented at 50 db (SPL) and a different button if it was presented at 100 db (SPL). After either twenty presentations of each stimulus, with reinforcement of correct responding, or after 20 correct responses in a row (whichever came first), additional intensity levels were inserted in 5 db steps between the two training stimuli. Each intensity level was presented 10 times during testing for a total of 110 stimulus presentations. Reinforcement of responding was withheld and Ss were instructed to respond to each stimulus. Figure 7 (top graph) shows (for the pooled data of 20 Ss the relative frequency of each response to the indicated stimulus intensities. The cross-over point of the two functions (the point of equal response probability) occurs at approximately 84.7 db SPL. The midpoint of the sone scale for the 50 to 100 phon range is 90.5 db SPL. The decibel midpoint is 75 db. The weighted mean of the sone scale values for all intensity levels (weighted according to the number of times each level occurred in the experiment, i.e., 10 times for each level except the training intensities which appeared on the average for different Ss a total of 25 times each) is, in phons, computed to be 84.8 db. (this prediction accords well with the obtained result: 84.7 db). The mean value predicted without differentially weighting the training stimuli is 82.8 db.

In a second part of the experiment described above, another group of 20 Ss were given discrimination training with three stimulus intensities, 50, 75, and 100 db SPL, which were correlated with reinforcement for three separate responses, R_1, R_2 and R_3, respectively. Each stimulus was presented during training at least 15 times. The Ss were then given the same generalization test for three responses that their counterparts received for two responses. The relative frequency of each of the three responses to the indicated stimulus intensities is shown in Fig. 7 (bottom graph). It is noteworthy that the stimulus maximally evoking R_2 during generalization testing was not the original training stimulus, 75 db. Stimulus control was transferred to the right on the continuum, to a point closer to the mean value of the stimuli involved. The displaced peak was predicted to be located at 84.1 db, taking into consideration the differential weights assigned to the three training stimuli. The results of these two experiments provide evidence for the relativity of the controlling relations obtaining between stimuli and responding, and indirect support for the view that the mean value arising in an experiment is affected by preceding levels of stimulation as well as prevailing levels.

Thus, a generalized concept of mean value, operating on psychological units derived from direct estimations of apparent magnitudes along some continuum, and taking relativity or AL effects into account by means of weightings, has proven initially to be an accurate predictor of two parameters of stimulus generalization functions along that continuum.

Effects of prior stimulation on mean value levels.

Two-alternative, forced-response generalization measures were obtained for pure-tone stimuli varying in frequency (cps), following discrimination training with different pairs of tones drawn from this continuum. Pure tones with frequencies of 200, 300, 400, 600, 800, 1200, 1600, and 2400 cps, and duration 400 msec, were presented to S binaurally with a sound pressure level of 95 db (as measured at the TDH-39 earphones). The stimuli were produced by a push-button audio oscillator (Hewlett-Packard 241A) and pre-recorded on magnetic tape. The stimuli were measured with a frequency counter (Hewlett-Packard 523DR) and found always to be within less than .5 percent of their nominal values.

Thirty undergraduate students served individually as Ss in sessions lasting approximately 30 min. The Ss were assigned randomly to three groups of 10 each. Each group was given the same generalization test, comprising 16 presentations of each of the 8 stimulus levels randomized by blocks. Before administering this test, however, the Ss were trained to respond differentially by button press to a pair of tones presented 15 times each. The pair of tones used differed for the 3 groups. Group I received training with S_1 = 200 cps and S_2 = 300 cps. Group II Ss were trained with S_1 = 1600 cps and S_2 = 2400 cps. Group III Ss received training with S_1 = 200 cps and S_2 = 2400 cps.

The relative frequency of R_1 to each tone in generalization testing is shown in Fig. 8. These curves represent the pooled data of 8 Ss from group I, and 9 Ss from groups II and III respectively. It had been decided beforehand to reject the data from all Ss who failed to give monotone response functions. The response functions of the 4 Ss rejected had large inversions of the first order. The results for the individual Ss remaining, within a group, were very similar. An analysis of variance performed on the points of equal response probability, estimated for each S individually by linear interpolation on the log frequency scale, resulted in an F-ratio = 10.06 for the group effect which, with 2 and 24 degrees of freedom, is significant at the .001 level.

Characterization of mean values for arbitrary functions.

The generalized means $M_t(x)$ are of the form

$$M_\phi(x) = \phi^{-1}\left[\sum_{i=1}^{n} q_i \phi(x_i)\right] \tag{12}$$

where $\phi(x)$ is one of the functions x^t, or log x and $\phi^{-1}(x)$ is the inverse function. Although these two functions are the most commonly considered in formulating psychophysical laws it is natural to consider the more general means formed from arbitrary functions. Aczél (1948) presented the necessary and sufficient conditions characterizing mean values in general for any number of variables.

Representation theorem for nonsymmetric means. If the single-valued function $M_\phi(x_1, \ldots, x_n)$ of n variables, $a \leq x_i \leq b$, $(i = 1, 2, \ldots, n)$, satisfies the following conditions:

(i) reflexivity: $M_\phi(x,\ldots,x) = x$,

(ii) strict monotonicity: if $x_k \gtrless x'_k$ then

$$M_\phi(x_1, \ldots, x_k, \ldots, x_n) \gtrless M_\phi(x_1, \ldots, x'_k, \ldots x_n),$$

(iii) continuity:

$$|M_\phi(\bar{x}_1, \ldots \bar{x}_n) - M_\phi(x_1, \ldots, x_n)| < \varepsilon \text{ if } |\bar{x}_i - x_i| < \delta, \ (i = 1, 2, \ldots, n),$$

(iv) bisymmetry: the symmetry of the following function in its n^2 variables

$$\psi(x_{11}, \ldots, x_{1n}; \ldots; x_{n1}, \ldots, x_{nn}) = M_\phi[M_\phi(x_{11}, \ldots, x_{1n}), \ldots, N_\phi(x_{n1}, \ldots, x_{nn})],$$

then and only then there exists a continuous function $\phi(x)$ and real numbers $(q_1, \ldots, q_n)$, $q_i > 0$, $q_1 + \ldots + q_n = 1$, for which $M_\phi(x)$ has the form (12).

Four properties of mean values for arbitrary functions. For a discussion see Hardy, Littlewood, and Polya (1959); the properties are stated below without proof:

Interness. From (i) and (ii) above it follows that if $a < x_i < b$ for all x_i $(i = 1, 2, \ldots, n)$ then $a < M_\phi(x) < b$.

Uniqueness. In order that $M_\phi(x) = M_\psi(x)$ for all x_i and q_i it is necessary and sufficient that $\phi = \alpha\psi + \beta$, ($\alpha$ and β constants, $\alpha \neq 0$).

The following two properties limit the class of functions satisfying (1 :

A. For all positive x, q, and k if $M_\phi(kx) = kM_\phi(x)$ then and only then M_ϕ is M_t. In other words M_t are the only homogeneous means M_ϕ. This means that

$$\phi = \begin{cases} \alpha \log x + \beta \\ \alpha x^t + \beta \end{cases}$$

B. For all positive x, q, and k, if $M\ (x + k) = M\ (x) + k$ then and only then

$$\phi = \begin{cases} \alpha x + \beta \\ \alpha\lambda^x + \beta, \quad (\alpha > 0,\ \lambda > 1 \text{ or } \alpha < 0,\ 0 < \lambda < 1). \end{cases}$$

References

Aczél, J. On mean values. *Bull. Amer. Math. Soc.*, 1948, 54, 392-400.

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. A new phenomenon in time judgment. *Nature*, 1953, 172, 901-903.

Coombs, C. H. *A theory of data*. New York: Wiley, 1964.

Ferguson, A. *et al*. Quantitative estimates of sensory events. (A. Ferguson, C. S. Myers, R. J. Bartlett, H. Barrister, F. C. Bartlett, Wm. Brown, N. R. Campbell, K. J. Craik, J. Drever, J. Guild, R. A. Houstoun, J. D. Irwin, G. W. C. Kaye, S. J. F. Philpott, L. F. Richardson, J. H. Shaxby, T. Smith, R. H. Thouless, W. S. Tucker) *Advancement of Science*, 1939-40, 1, 331-349.

Gage, F. H. The measurability of auditory sensations. *Proc. Royal Soc.*, 1934, 116b, 103-109.

Garner, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1954, 26, 73-88.

Goode, Frank M. An algorithm for the additive conjoint measurement of finite data matrices. Paper presented at the Amer. Psychol. Assn. meetings, Los Angeles, September, 1964.

Hardy, G. H., Littlewood, J. E., & Polya, G. *Inequalities*. Cambridge: Cambridge Univ. Press, 1959.

Helson, H. *Adaptation-level theory: an experimental and systematic approach to behavior.* New York: Harper & Row, 1964

Hölder, O. Die axiome der quantität und die lehre vom mass. *Ber. Sach., Gesellsch. Wiss., Math.-Phys. Klasse*, 1901, 35, 1-64.

Kopp, J. L., & Cross, D. V. A relation between stimulus generalization and magnitude estimation. Unpublished report, 05613-3-P, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, Michigan, 1964.

Lewis, D. Quantitative methods in psychology. New York: McGraw-Hill, 1960.

Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: a new type of fundamental measurement. J. math. Psychol., 1964, 1, 1-27.

Pfanzagl, J. A general theory of measurement applications to utility. Naval Res. Logistics Quart., 1959, 6, 283-294. (a)

Pfanzagl, J. Die axiomatischen grundlagen einer allgemeinen theorie des messens. Schrift. Stat. Inst. Univ. Wien, Neue Folge Nr. 1, 1959. (b)

Restle, F. Psychology of judgment and choice: a theoretical essay. New York: Wiley, 1961.

Ross, S. Logical foundations of psychological measurement. Copenhagen: Munsgaard, 1964.

Stevens, S. S. A scale for the measurement of a psychological magnitude: loudness. Psychol. Rev., 1936, 43, 405-416.

Stevens, S. S. The measurement of loudness. J. acoust. Soc. Amer., 1955, 27, 815-829.

Stevens, S. S. On the psychophysical law. Psychol. Rev., 1957, 64, 153-181.

Stevens, S. S. The psychophysics of sensory function. Amer. Scient., 1960, 48, 226-253.

Stevens, S. S., & Volkman, J. The relation of pitch to frequency: a revised scale. Amer. J. Psychol., 1940, 53, 329-353.

Suppes, P., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology. New York: Wiley, 1963.

Table 1

Intensity (in db SPL) of obtained bisection points.

Subject No. 1

Session

	I	II	III	IV
B(65, 75):	72.3	74.4	74.0	74.3
B(80, 90):	87.3	87.0	87.4	87.6
B(65, 80):	74.6	78.6	77.3	76.8
B(75, 90):	84.5	86.8	86.3	87.1
B[B(65, 75), B(80, 90)]:	83.0	84.0	83.5	83.7
B[B(65, 80), B(75, 90)]:	81.5	84.5	83.7	83.5

	V	VI	VII	VIII
B(75, 65):	70.9	70.4	72.3	72.2
B(90, 80):	86.1	86.3	85.5	85.5
B(80, 65):	72.9	74.0	75.6	75.9
B(90, 75):	81.3	83.5	84.4	81.7
B[B(90, 80), B(75, 65)]:	78.3	78.5	79.5	79.8
B[B(90, 75), B(80, 65)]:	79.0	79.0	80.0	79.9

Table II

Bisections of Pure Tone Intensity

Measured in db SPL

A. Subject No. 1

		S_b						
	S_x	60	65	70	75	80	85	90
	60	61.0	64.0	65.7	70.2	74.2	77.1	78.8
	65	63.9	66.0	68.5	71.6	74.3	76.6	80.0
	70	64.3	68.0	70.3	74.0	76.0	77.8	82.5
S_a	75	69.5	71.1	72.1	75.4	78.4	81.4	84.7
	80	70.3	74.0	75.0	77.5	80.1	83.6	86.5
	85	77.6	77.0	79.1	82.4	83.5	84.9	88.1
	90	77.9	82.6	84.2	85.6	86.5	87.3	89.9

B. Subject No. 3

		S_b						
	S_x	60	65	70	75	80	85	90
	60	61.0	61.8	64.5	66.3	69.2	73.5	77.8
	65	62.9	65.2	67.3	69.5	71.5	75.4	80.5
	70	65.6	67.6	70.3	73.0	75.4	77.9	82.6
S_a	75	68.0	69.4	71.9	74.6	77.3	80.7	84.8
	80	71.6	73.6	75.0	77.4	79.6	82.5	86.5
	85	74.2	75.4	77.5	81.0	82.4	84.9	88.0
	90	76.9	77.8	81.4	84.0	85.8	88.5	89.8

Figure Captions

Fig. 1. Diagram of the apparatus involved in the bisection experiments.

Fig. 2. Bisection intensity *x* as a function of the intensity of the first tone *a* (abscissas) and last tone *b* (parameter) in the sequence. Data for *S*3.

Fig. 3. Bisection intensity *x* as a function of the intensity of the first tone *a* (abscissas) and last tone *b* (parameter) in the sequence. Data for *S* 1.

Fig. 4. Computed values for $f(x) = \frac{1}{2} \left[f(a) + f(b)\right]$ from the data of *S*3 as a function of obtained bisection adjustments in phons.

Fig. 5. Log f(x) computed from data of *S*1 as a function of obtained bisection adjustments in phons.

Fig. 6 Relation between a constructed loudness scale, derived from the bisections by *S*1, (with the origin estimated from Fig. 5) and loudness level in phons.

Fig. 7. Generalization following discrimination training with two stimuli (top graph) and three stimuli (bottom graph).

Fig. 8. Gradients of generalization to pure tone frequency as a function of the stimuli used in discrimination training.

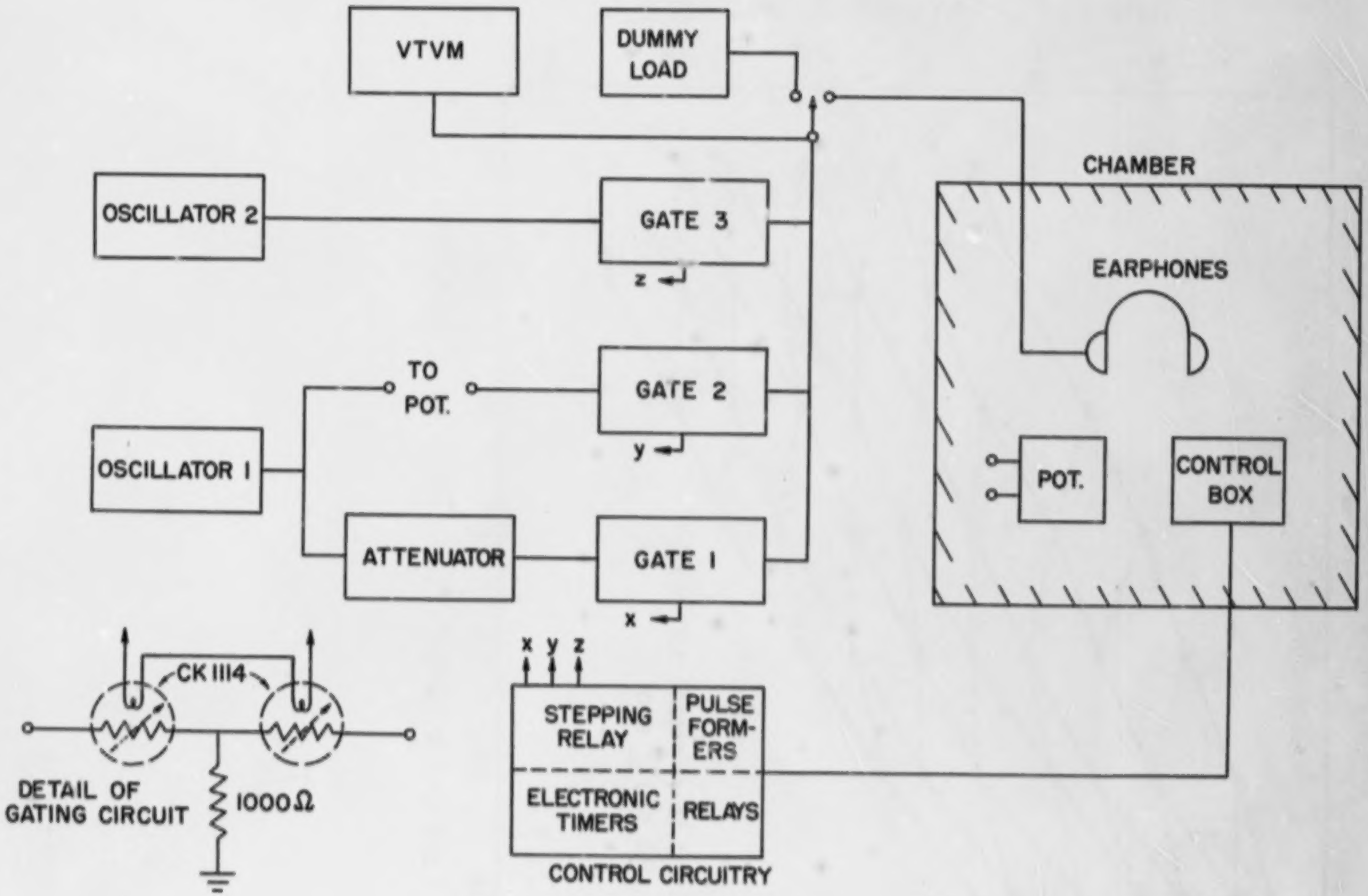

VTVM
DUMMY LOAD
CHAMBER
OSCILLATOR 2
GATE 3
z
EARPHONES
TO POT.
GATE 2
y
OSCILLATOR I
POT.
CONTROL BOX
ATTENUATOR
GATE I
x
x y z
CK 1114
STEPPING RELAY
PULSE FORMERS
ELECTRONIC TIMERS
RELAYS
CONTROL CIRCUITRY
DETAIL OF GATING CIRCUIT
1000Ω

Fig. 1

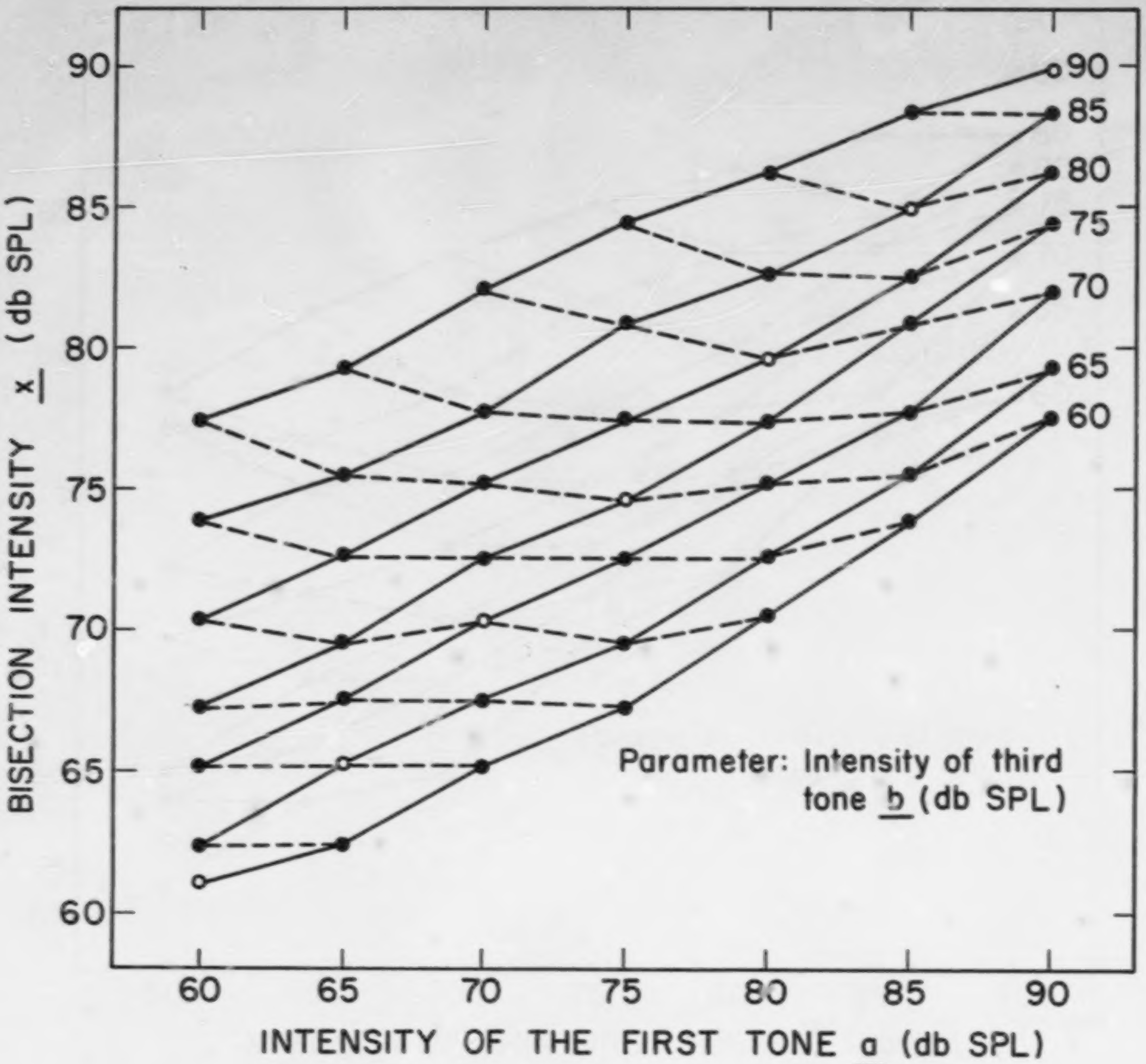
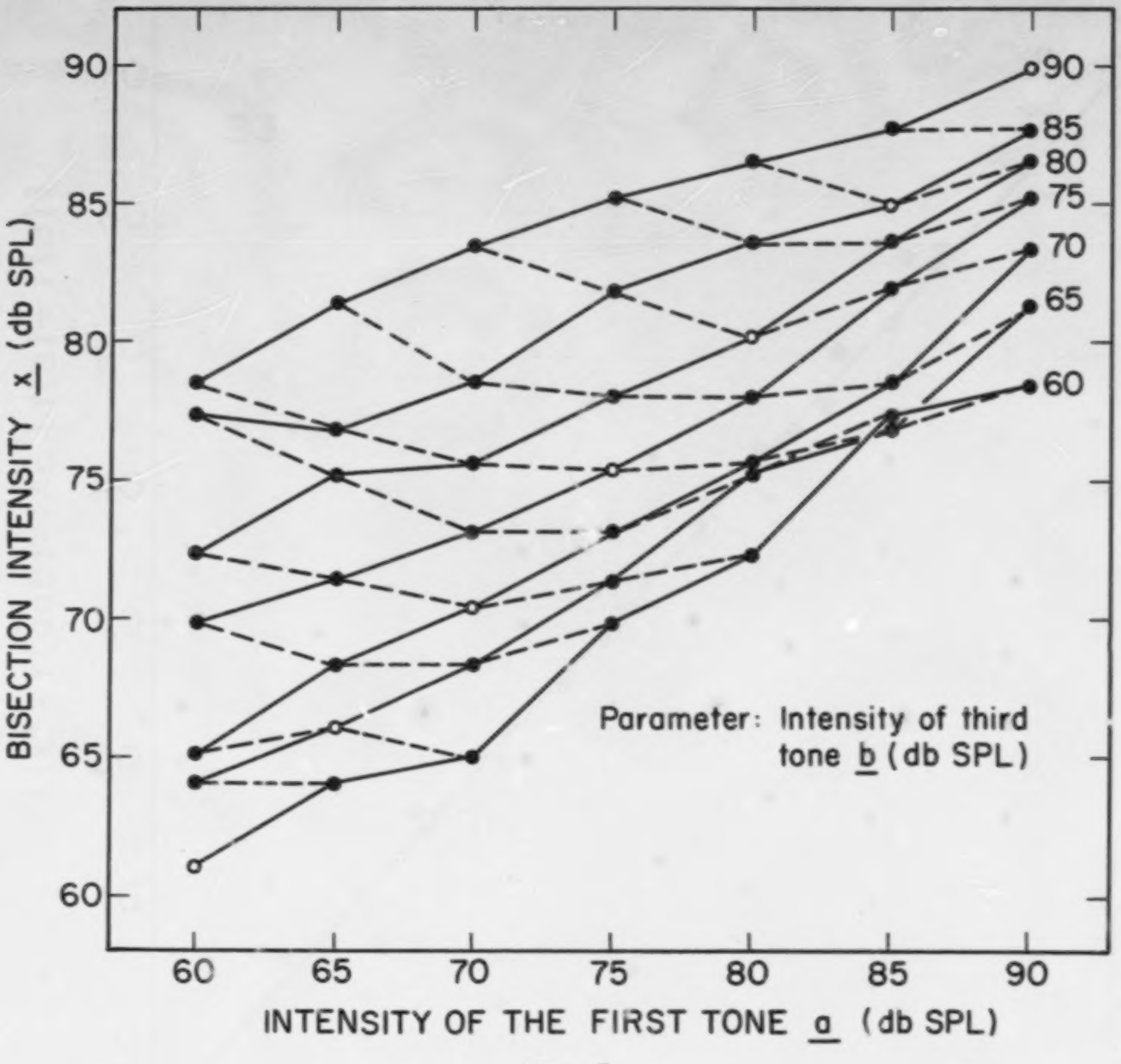

BISECTION INTENSITY x (db SPL)
90
85
80
75
70
65
60
INTENSITY OF THE FIRST TONE a (db SPL)
60
65
70
75
80
85
90
Parameter: Intensity of third tone b (db SPL)
90
85
80
75
70
65
60

Fig. 2



90
85
80
75
70
65
60
BISECTION INTENSITY x (db SPL)
60
65
70
75
80
85
90
INTENSITY OF THE FIRST TONE a (db SPL)
90
85
80
75
70
65
60
Parameter: Intensity of third tone b (db SPL)

Fig. 3

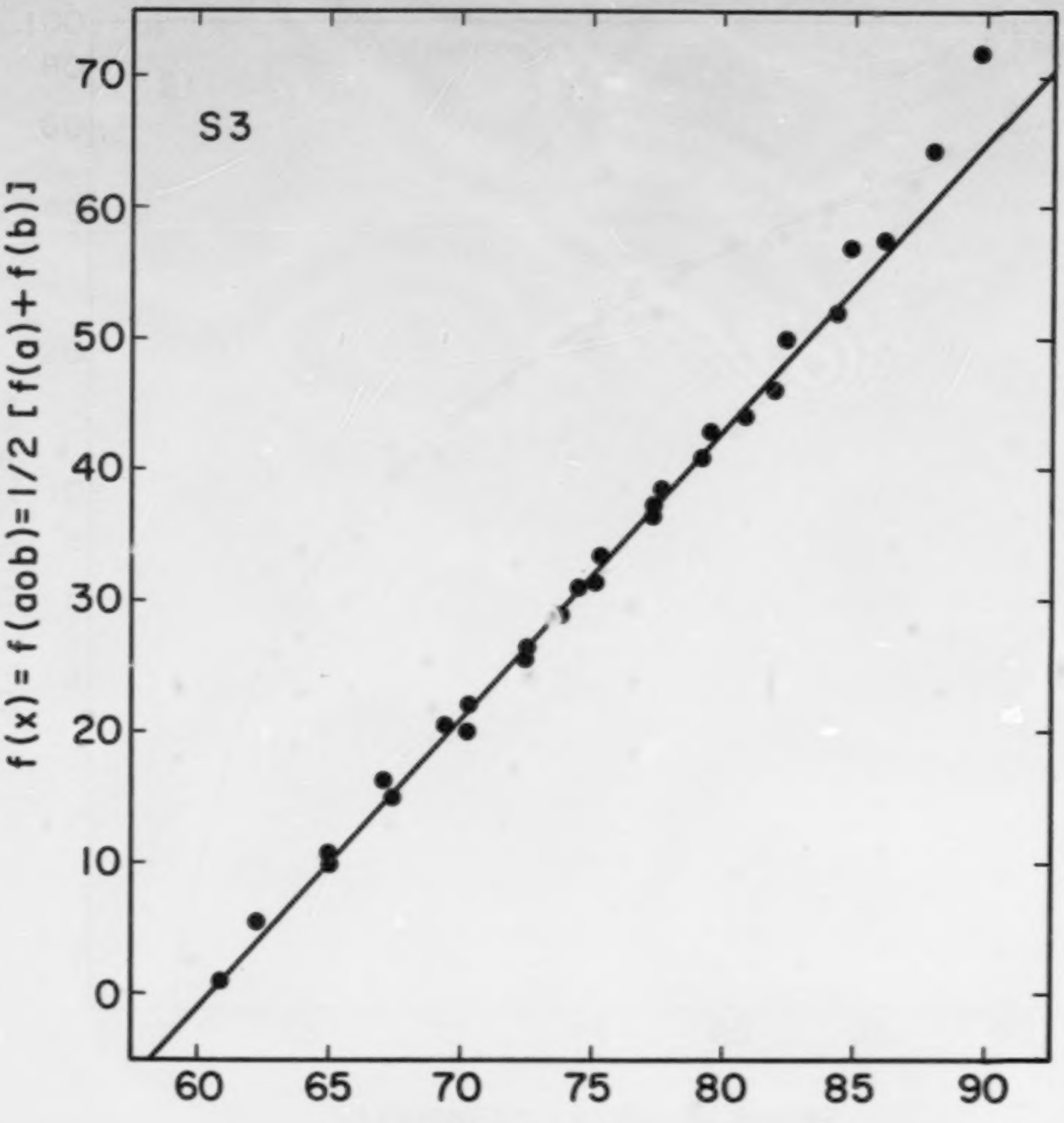

S3
f(x) = f(aob) = 1/2 [f(a) + f(b)]
0
10
20
30
40
50
60
70
60
65
70
75
80
85
90
LOUDNESS LEVEL IN PHONS

Fig. 4

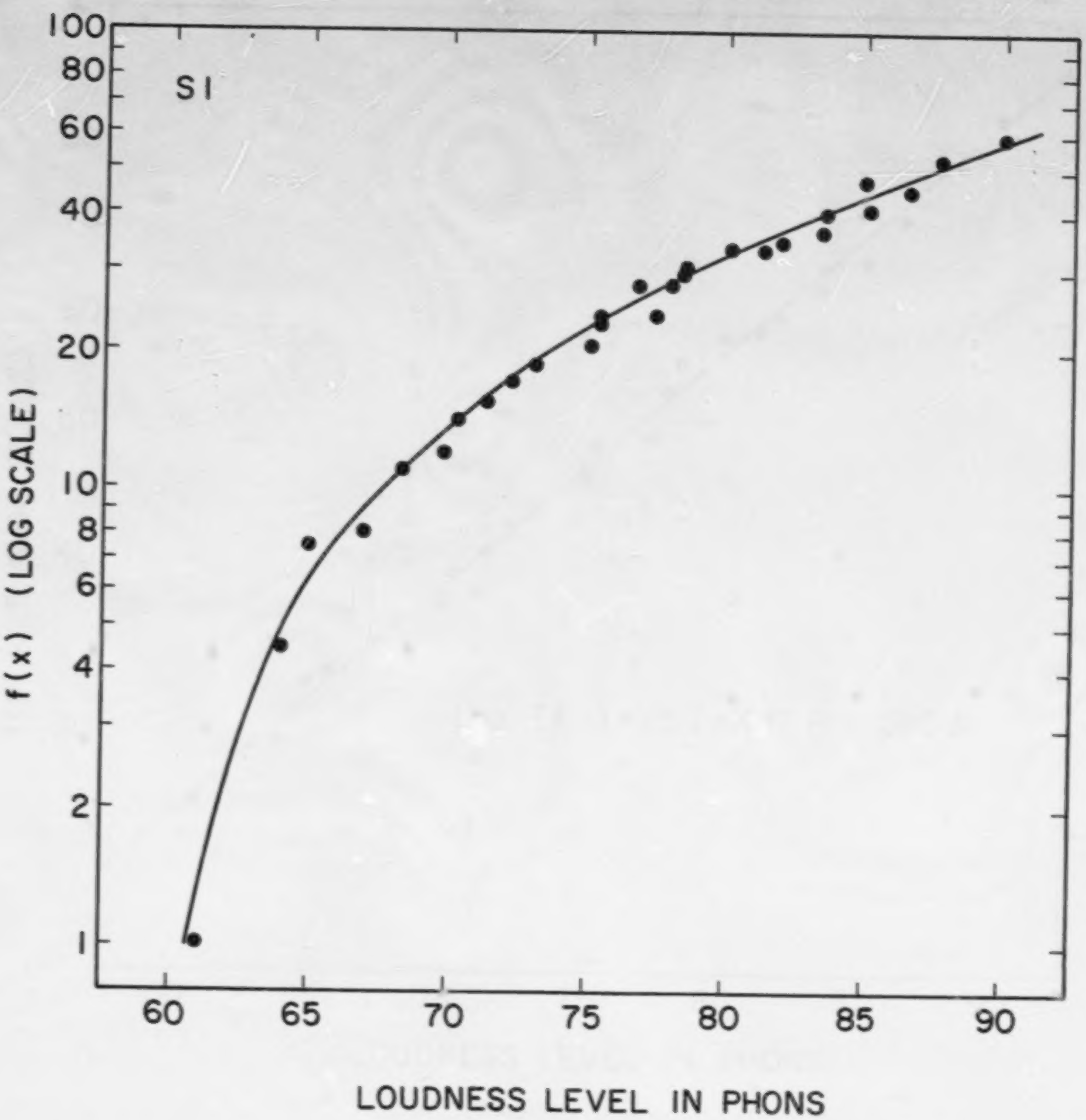

S1
f(x) (LOG SCALE)
100
80
60
40
20
10
8
6
4
2
1
60
65
70
75
80
85
90
LOUDNESS LEVEL IN PHONS

Fig. 5

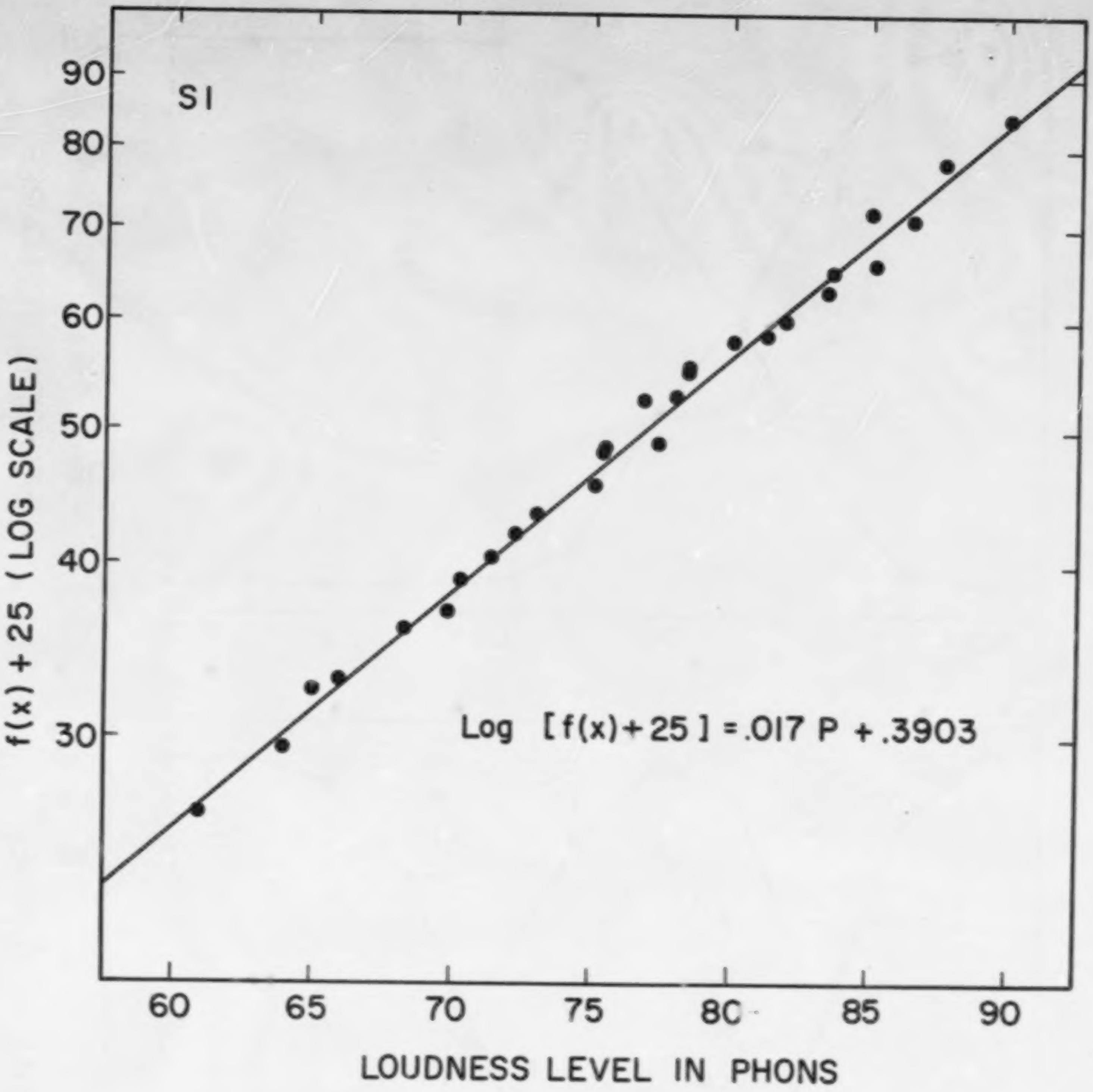

S1
Log [f(x)+25] =.017 P +.3903
f(x) + 25 (LOG SCALE)
90
80
70
60
50
40
30
60
65
70
75
80
85
90
LOUDNESS LEVEL IN PHONS

Fig. 6

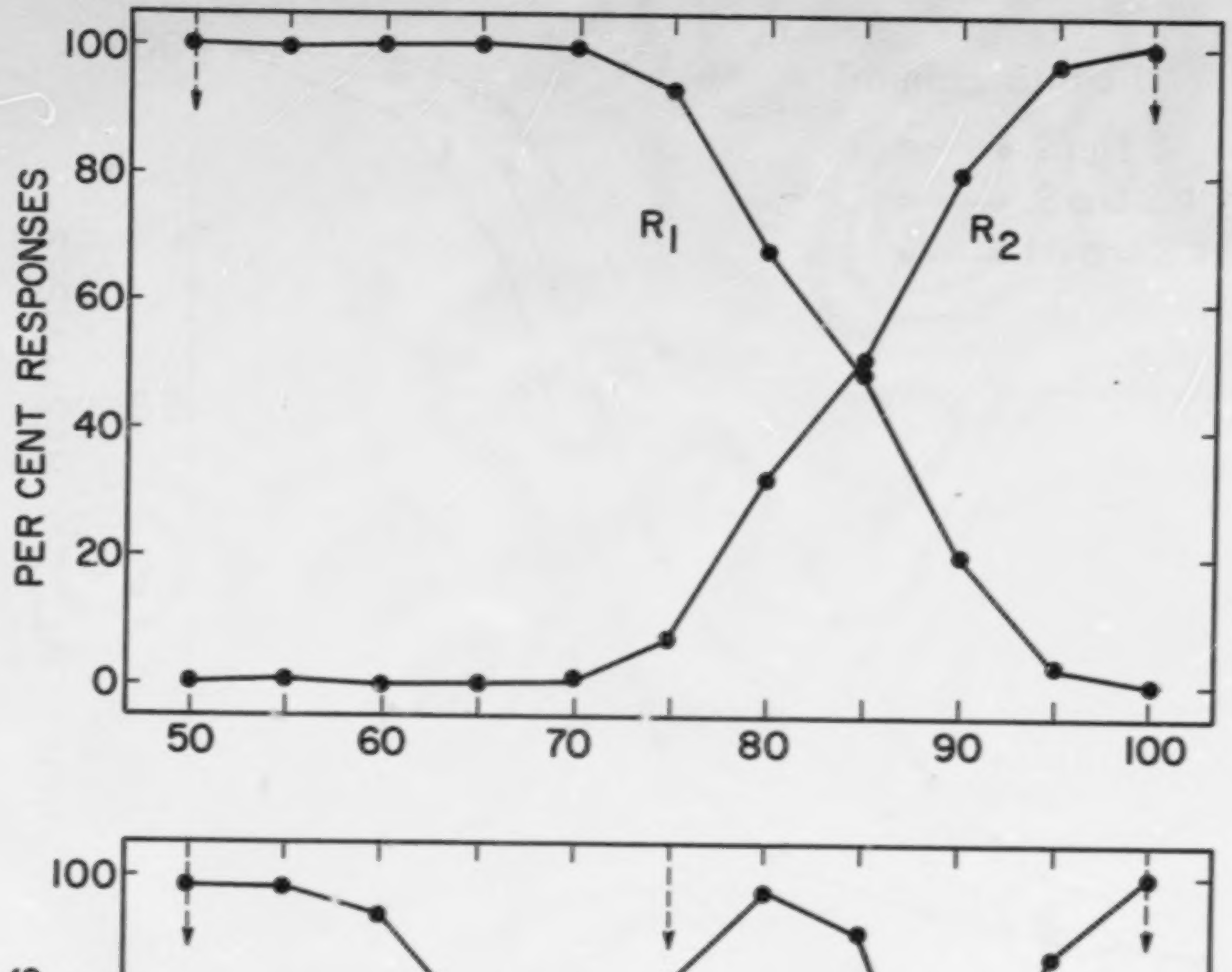

PER CENT RESPONSES
0
20
40
60
80
100
R1
R2
R3
50
60
70
80
90
100
TONE INTENSITY (db SPL)

Fig. 7

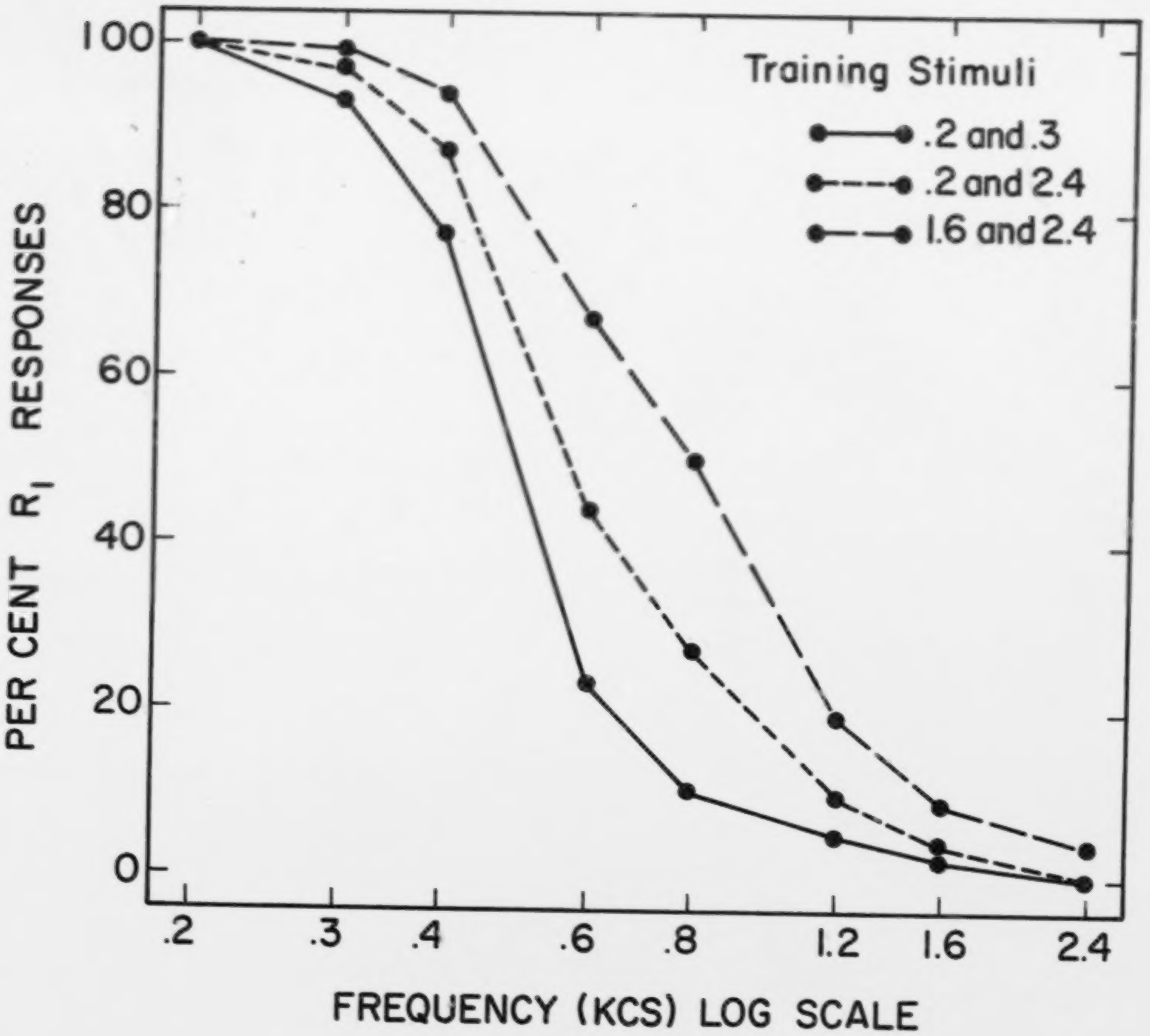

100
80
60
40
20
0
PER CENT R_1 RESPONSES
.2
.3
.4
.6
.8
1.2
1.6
2.4
FREQUENCY (KCS) LOG SCALE
Training Stimuli
.2 and .3
.2 and 2.4
1.6 and 2.4

Fig. 8

A SELF-INSTRUCTIONAL DEVICE FOR CONDITIONING ACCURATE PROSODY

Roger Buiten and Harlan Lane
The University of Michigan

A Self-instructional Device
for Conditioning Accurate Prosody [1]

Roger Buiten and Harlan Lane
Behavior Analysis Laboratory
The University of Michigan

The achievement of native-like fluency in a second language is as uncommon and arduous as it is essential to effective social behavior in a foreign land. Social custom generally accords the novice who is acquiring a set of skills--whether in sports, cooking, warfare, or education--both indulgence and encouragement, but this is not the fare of a novice speaker of a second language. Native speakers of the language view him with suspicion at least, and at worst with ridicule and hostility. Perhaps this is because of the central importance of language in almost every aspect of social life. Indeed, philosophers have contended that language is the defining property of man; the layman's tendency to view his own language as the defining property of man may have the same sources. Mark Twain entertainingly reveals the vital social role of fluency in the language of the community in this dialogue between Huck Finn and his friend Jim:

> "Looky here, Jim; ain't it natural and right for a cat and a cow to talk different from us ?"

---

[1] An address to the International Congress of Applied Psychology August, 1964.

> "Why, mos' sholy it is."
> "Well, then, why ain't it natural and right for a Frenchman to talk different from us? You answer me that."
> "Is a cat a man, Huck?"
> "No."
> "Well, den, dey ain't no sense in a cat talkin' like a man. Is a cow a man?--er is a cow a cat?"
> "No, she ain't either of them."
> "Well, den she ain't got no business to talk like either one er the yuther of 'em. Is a Frenchman a man?"
> "Yes."
> "Well, den! Dad blame it, why doan' he talk like a man? You answer me dat!"

What can be done to teach a Frenchman, or an American, or a Yugoslav to "talk like a man" when he is among men who talk in a language different from his own? Clearly, the speaker must master the component speech sounds of the second language, their permissible patterns in grammatical utterances, and the appropriate occasions for these utterances. Without these complex skills he will often be unintelligible (Lane 1963a), and so it is these skills that receive the lion's share of research and pedagogy. However, intelligibility is only a part of the story of fluency in a foreign language. This report of research is concerned with another set of skills, relatively independent of the first, which contributes little to intelligibility and everything to fluency, and thus to social acceptance. Native-like fluency in a second language is predicated on a mastery of its patterns of intonation, stress, and rhythmn -- in short, a mastery of its prosody.

Phonetic descriptions of numerous and diverse languages reveal a variety of highly stylized patterns of stress and intonation. The novice speaker of a second language who employs the prosody of his native tongue is most likely to sound foolish, even if his first and second languages have common historical roots and belong to the same language family.

Recent acoustic studies of language prosody [2] confirm and amplify the phonetic descriptions. Again, the theme is diversity between languages and stereotypy within them. Delattre (1962a), for example, has extracted the characteristic intonation contours for declarative sentences shown in Fig. 1 from spectrographic analyses of conversations in four languages. Among other contrasts,

insert Fig. 1 about here

American intonation seems to be characterized by a recurrence of falling contours, whereas the motif of German, Spanish and French is the rising contour. Similarly, there are characteristically different patterns of stress among languages. Concerning these four languages, for example, Delattre provides evidence that in French the stress is characteristically on the final syllable, whereas in Spanish it falls on the next-to-last syllable, while English and German vary between these two loci. The temporal patterns of responding, or rhythmic characteristics of diverse languages, have received little study, but the covariance of duration and stress in several languages is well documented (e.g., Lieberman, 1960), and it is not unlikely that future research will also reveal stylized patterns in the temporal spacing of speech segments.

---

[2] Dreher (1951), Goodman (1952), Fry (1955, 1958), Malmberg (1956), Bolinger (1958a, 1958b), Bascom (1959), Jassem (1959), Garding & Gerstman (1960), Lehiste (1960, 1961), Lieberman (1960), Delattre (1961, 1962a, 1962b), Goldman - Eisler (1961), Hadding - Koch (1961), Rigault (1962), Lindblom (1963), Magdics (1963)

It is axiomatic for the present report, then, that languages differ in their prosody and that the speaker who has not mastered the prosody of a second language is not fluent in that language. Before describing our approach to conditioning prosodic skills in a second language, it is desirable to briefly review the alternative approaches that have been employed--largely unsuccessfully--in the past. We restrict our attention to teaching prosodic accuracy to an adolescent or adult while he resides in his native land, with all the sources of resistance to the learning of new habits of speech that this implies (Lane, 1962a). First, and most common, there are the techniques available to the language teacher in the traditional classroom setting. In this case, prosodic skills literally go by the board; they are rarely acquired well. Several constraints on language learning in the classroom militate against the acquisition of prosodic skills: the widespread conviction among college students that a language block is a place of execution, the broad objectives of the language course, often ranging from culture through literature to spelling, the large number of students confronting each teacher, the eddy of second-language behaviors created amidst a sea of first-language habits, and so forth (Lane 1964a; 1964b). Rarely are the avowed ingredients of learning in evidence: active responding by the individual student, differential feedback or "reinforcement" depending of the accuracy of his responses, and careful sequencing of each component task according to the student's individual level of proficiency at the moment (Lane, 1963b). Moreover, the intrinsic complexity of prosodic features, extending as they do over several segments of an utterance, presents a special measure of difficulty to their attainment by classroom instruction. Little wonder if the language teacher is pleased when in two or three years a student learns the segmental sounds, some vocabulary, and the elements of reading and writing in the second language. If he doesn't sound like an authentic Frenchman, well, after all, he's not one.

When choral work in the language classroom is supplemented by sonic treatments in the language laboratory, the gain in learning is small.[3] Simple exposure is no more effective than it is plausible as a basis for learning (Lane, 1962b). In a few language laboratories there is provision for active responding by the student and in a lesser number he may even hear his poor pronunciation played back to him repeatedly. The student is typically in no position to evaluate his pronunciation, however, since the novice speaker is usually a novice listener. Even if the student were able to discriminate between good and poor pronunciation, these have no differential effects in the immediate environment. When audible differences don't make a difference, both discrimination and production fail.

The third and most recent approach to language pedagogy is programmed instruction (Lane, 1964a). The application of programmed instruction has necessarily involved a molecular functional and linguistic analysis of the behaviors involved in language learning. Based on this analysis, new techniques for controlling the learning process have been applied to rendering it more effective and efficient. Once again, however, the conditioning of prosodic features has largely been overlooked. None of the language programs

---

[3] Lewis (1961), Reichard (1962), Keating (1963), Hutchinson & Gaarder (1964), Scherer (1964).

described in several recent and comprehensive surveys explicitly states prosodic accuracy as an objective, nor does any of these programs explicitly provide for the conditioning of the prosodic features, although all, of course, do battle with the segmental features and with vocabulary. The obstacles to employing programmed self-instruction as a means for conditioning prosodic accuracy are these. The student must be taught to discriminate between accurate and inaccurate prosody in the first place; then, these discriminations must be maintained during the acquisition of productive skills; finally, the student must apply the discriminations to his own vocal behavior--reliably and validly. No language program has yet tackled, no less has overcome, these obstacles.

From the ashes of these unsuccessful methods for conditioning prosodic accuracy in a second language, arises the image of a pedagogical phoenix with the following plumage. An instructional method is required that will: 1) *provide the student with rapid, reliable, and valid evaluations of his prosodic accuracy*, and 2) *adapt the learning sequence in accord with the student's proficiency at each step along the way*. These pedagogical requirements led to the development, at the Behavior Analysis Laboratory, of a device called SAID, the Speech Auto-Instructional Device.

SAID is an electro-mechanical device which performs three significant functions in conditioning prosodic accuracy in a second language.[4] First, SAID presents to the student tape-recorded pattern sentences that are considered standards in prosodic performance. These sentences are programmed in the best known sequence for teaching prosody in the target language to a speaker of a given native language. The student is instructed to imitate the pattern sentence after he hears it. Second, SAID processes the student's imitation, and instantaneously evaluates its acceptability on the basis of its three distinct prosodic features: pitch, loudness, and tempo. Third, SAID immediately displays to the student the degree to which his imitation is unacceptable, and demonstrates how he must modify his next imitation, in the prosodic feature under consideration, in order to make the imitation more acceptable. This process of presentation-evaluation-display repeats itself until the prosody of the student's imitation is acceptable.

---

[4] A more detailed and technical account of the operation of the SAID system will be found in Lane and Buiten (1964).

The first phase of the three-fold machine process is the presentation of the tape-recorded pattern sentences. The sequence of patterns presented to the student is controlled by the computer base of SAID (Fig.2), which instructs the tape recorder to rewind when the imitation is unacceptable or to play the next pattern when the imitation is accurate.

---

insert Fig. 2 about here

The second phase, evaluation, is accomplished both by the computer and by analog electronic equipment. The speech signals from the tape recorder and the student's response microphone are channelled to analog extractors of the acoustic parameters (Fig. 3). The conceptual bases for the parameter extractors were contributed by the University of Michigan's Communication

insert Fig. 3 about here

Sciences Laboratory. The fundamental frequency, the major contributor to the perception of pitch in speech (Lane, 1962c), is extracted by filtering each of the speech signals into two ranges within the entire fundamental-frequency spectrum of the human voice. The range containing the fundamental is selected by switching circuitry and the signal from the selected range is sent to a frequency-to-voltage converter. The output of the converter is a dc voltage that varies linearly with the fundamental frequency of the speech signal. In this form it is digitized and processed by the computer.

The intensity of the speech signal, the major contributor to the perception of loudness (Lane, Catania & Stevens, 1961), is extracted by rectifying and filtering. This produces a dc voltage which varies linearly with the peak instantaneous voltage of the speech signals. This dc voltage is digitized and processed by the computer. While processing the intensity parameter, the computer also records the time elapsed between successive intensity peaks.

The student has not been trained, and therefore is not able, to accurately interpret changes in frequency, intensity, and temporal spacing, since changes in these physical parameters do not produce simply corresponding changes in their apparent or psychological values. The student can, however, be expected to discriminate readily variations in the psychological effects of these physical parameters: pitch, loudness, and tempo. Therefore, the computer converts the speech parameters from linear physical units to logarithmic psychological units according to the psychophysical functions for speech production and perception determined in earlier research (Lane, 1962c). The variation of these psychological values more accurately represents the prosodic features of speech from the point of view of the student and his language community. While the pattern sentence is presented to the student, the values of its psychological parameters are stored away in the memory of the computer. While the student is imitating the pattern, the psychological parameter values of his imitation are compared to corresponding stored values of the pattern, and three error scores are generated. The error scores are compared to preset criteria of acceptability for the three features. The results of these acceptance tests determine the sequence of presentation and display.

The third phase of the machine process involves displaying the results of the comparison of pattern and imitation, and the results of the acceptance tests. The display of the comparison difference is made on a zero-center meter, so that both the direction and magnitude of error score can be monitored readily. The student's error score for one of the three features is displayed to him as his imitation proceeds (Fig. 4). A neon indicator on his panel shows him which feature is being displayed on each trial. When the student's imitation is acceptable -- that is, within preset tolerance -- in pitch, for example, the device advances to the next feature, loudness, and the cycle of presentation - evaluation - display is repeated for this feature. Ultimately, the imitation

---

insert Fig. 4 about here

---

is acceptable with respect to all three features, whereupon the next pattern sentence in the program is presented.

To illustrate the operation of the SAID system we may follow the student through one complete learning cycle. The student, seated in the sound-insulated booth shown in Fig. 4 above, listens to the presentation of the pattern sentence. A light on his indicator panel tells him to imitate the "pitch" contour of the pattern. While he is making his initation, the meter on his panel is swinging to the right when his response is too high in pitch, and to the left when his response is too low in pitch. If his largest error indication is greater than the tolerance for accurate imitation, the device will rewind the tape recorder and present the pattern sentence again. Assuming that the student's imitation was acceptable on this second attempt, the device turns on the "loudness" light, rewinds the tape recorder, and again plays the pattern sentence. SAID then repeats the sequence for loudness as for pitch, and for tempo when loudness is acceptably imitated. Finally, the three features must be acceptably imitated concurrently on three successive trials in order for the device to advance the student to the next pattern sentence in the program.

The criteria for accurate imitation have been set at less than 6 per cent error for pitch, 20 per cent error for loudness, and 10 per cent error for tempo. These criteria were selected initially on the bases of acoustic-phonetic research and preliminary experiments with SAID. Ultimately, the choice of criteria must come from systematic listening experiments with native speakers of the target language.

The computer's teleprinter (Fig. 2) keeps a running record of the behavior of SAID and the student. When the cycle begins, the teleprinter types the word "pitch", indicating the feature which the student is trying to imitate. Just before the pattern sentence is played to the student, the teleprinter types the letter "P"; the letter "S" is typed just before the device processes the student's imitation. After the imitation, the result of each comparison of pattern and imitation is typed out. If the student's imitation is acceptable, the teleprinter types the next feature, (in this case "loudness"); it then types the letter "P" for pattern again, and the sequence is repeated. When the imitation is thoroughly acceptable, the student is ready to move on to the next pattern in the program. The teleprinter then, types out a table of results for all three features for each of the preceding trials before advancing.

A graphic analysis of the learning that has taken place is readily prepared from these records. When pitch, loudness, or tempo differences between the pattern and the imitation are plotted as a function of successive trials, several kinds of trends may be observed; the most obvious is the rate of learning to effect control over the prosodic features. The student's consistency and short-term retention, and the relative difficulty of the prosodic features are also revealed by the plots of error versus trials. We use the term "consistency" to refer to the degree to which the error is steadily reduced prior to reaching criterion, and the term "retention" to refer to the degree to which low error performance is maintained thereafter.

Preliminary experiments are underway to evaluate the effectiveness of the SAID system in conditioning prosodic accurracy. In one experiment, tape recordings were prepared containing triads of synthetic vowels. Within each triad, the vowel stimuli varied in fundamental frequency, formant frequency, relative intensity, duration or temporal spacing. The error functions for a typical subject are shown for each of the three prosodic features in Figs. 5, 6 and 7. These data were obtained with a pattern of three vowels that varied only in fundamental frequency: 160, 120, and 160 cps. The stimulus duration was 250 msec; inter-stimulus duration was 500 msec; intensity was constant at about 70 db (SPL); the formant frequencies were 300 and 900 cps.

In the plot of pitch error versus successive trials (Fig 5), the filled squares represent those trials on which the student's error in pitch was displayed to him. This student, like several others, finds the task of imitating the pitch of the vowel pattern quite simple under the supervision of the SAID system. Within a few trials, he achieves and retains a level of accuracy that is well within tolerance.

---

insert Fig. 5 about here

The SAID system then advanced to loudness evaluation and error display. The student's performance with respect to this prosodic feature is shown in Fig. 6. The filled triangles indicate the error in relative loudness on those trials for which loudness error was displayed to the student. Although the student is much less consistent in imitating the loudnesses of the pattern than in imitating its pitch, continued attempts at imitation with visual display

---

insert Fig. 6 about here

---

of error soon bring the loudness feature of his prosody within the acceptance limits. It is noteworthy that there is an increase in loudness error when the student first turns his attention to imitating loudness. Retention seems to improve with training; the trend of the function approaches a point well below the acceptance limit.

Finally, the SAID system advanced to tempo evaluation and display. The student's performance is shown in Fig. 7, where the filled circles indicate the error in tempo on those trials for which tempo error was displayed. This student, like several others, is consistent in his imitations of tempo and quickly brings this feature of his prosody within tolerances, but his retention of accurate tempo is poor. The latter finding may be attributable in part to the fact that stimulus duration, inter-stimulus interval, and intensity distribution within each stimulus all may contribute to the perception of tempo. The student whose data are shown in Figs. 5-7 is typical in that tempo was easier to master than loudness, but more difficult to master than pitch.

---

insert Fig. 7 about here

Inspection of Fig. 7 reveals another programming feature that is presently a part of the SAID system. When the student achieves an acceptable imitation of tempo, the device tests the other two prosodic features concurrently. If the error associated with either feature is greater than criterion, the system recycles to evaluation and error display of the unacceptable feature. If both of the other features are within criterion, a light is flashed on the student's panel. This procedure is repeated until the student accurately imitates all three features on three successive trials. Then, the next pattern in the program is presented.

The capability of the SAID system to provide rapid, reliable and valid evaluations of prosodic accuracy and to adapt the learning sequence in accord with the student's proficiency at each step, as well as the findings of initial experiments with the system, hold promise of improving and accelerating an important facet of second-language learning. These capabilities and findings, however, raise certain preliminary experimental questions that are normally overlooked by traditional methods of language pedagogy. What are the quantitative tolerances of the prosodic features for native-like fluency within a language? What are the sources of facilitation and interference between prosodic skills in a given pair of native and target languages? How well are newly acquired prosodic skills retained outside the learning environment and what can be done to enhance this retention? These and related questions will be examined in future research with the Speech Auto-Instructional Device.

References

Bascom, J. Tonomechanics of Northern Tepuhan, Phonetica, 1959, 4, 77-88.

Bolinger, D. L. A theory of pitch accent in English. Word, 14, 1958, 109-149. (a)

Bolinger, D. L. On intensity as a qualitative improvement of pitch accent. Lingua, 1958, 7, 175-182. (b)

Delattre, P. La lecon d'intonation de Simone de Beauvoir, etude d'intonation declarative comparee. French Rev., 1961, 35, 59-67.

Delattre, P. The general phonetic characteristic of languages. Boulder, Colo., University of Colorado, Dept. of French, 1962. (a)

Delattre, P. A comparative study of declarative intonation in American English and Spanish. Hispania, 1962, 45, 233-241. (b)

Dreher, J. A comparision of native and acquired language intonation. Univer. of Michigan doctoral dissertation, 1951.

Fry, D. B. Duration and intensity as physical correlates of linguistic stress. J. acoust. Soc. Amer., 1955, 27, 765-768.

Fry, D. B. Experiments in the perception of stress. Lang. & Speech, 1958, 1, 126-152.

Garding, E. & Gerstman, L. J. The effect of changes in the location of an intonation peak on sentence stress. Studia Ling., 1960, 14, 57-59.

Goldman-Eisler, F. The significance of changes in the rate of articulation. Lang. & Speech, 1961, 4, 171-174.

Goodman, A. C. Imitation of intonation patterns. Univer. of Michigan doctoral dissertation, 1952.

Hadding-Koch, K. Acoustical-Phonetic Studies in the Intonation of Southern Swedish. Lund-Gleerup, 1961. (Travaux de l'institut de Phonetique de Lund III)

Hutchinson, J. C. & Gaarder, A. B. *A brief analysis of the Keating Report*. Unpublished report, Language Development Section. U. S. Office of Education, Washington, D. C., 1964.

Jassem, W. The phonology of Polish stress. *Word*, 1959, 15, 269-282.

Keating, R. F. *A study of the effectiveness of language laboratories*. New York: Inst. Admin. Res., Teachers College, Columbia Univer., 1963.

Lane, H. L. Some differences between first and second language learning. *Lang*. *Learn*., 1962, 12, 115-123. (a)

Lane, H. L. Experimentation in the language classroom: guidelines and suggested procedures for the classroom teacher. *Lang*. *Learn*., 1962, 12, 115-123. (b)

Lane, H. L. Psychophysical parameters of vowel perception. *Psychol*. *Monogr*., 1962, 761, Whole no. 563. (c)

Lane, H. L. Foreign accent and speech distortion. *J*. *acoust*. *Soc*. *Amer*., 1963, 35, 451-453. (a)

Lane, H. L. Specifications for auditory discrimination learning in the language laboratory. *Int*. *J*. *Amer*. *Ling*., 1963, 29, 61-69. (b)

Lane, H. L. Programmed learning of a second language. *Int*. *Rev*. *Applied Ling*., 1964 (in press). (a)

Lane, H. L. Acquisition and transfer in auditory discrimination. *Amer*. *J*. *Psychol*., 1964, 77, 240-248. (b)

Lane, H. & Buiten, R. *A preliminary manual for the speech auto-instructional device*. Unpublished report, Behavior Analysis Laboratory, Univer. of Michigan, 1964.

Lane, H. L., Catania, A. C. & Stevens, S. S. Voice level: autophonic scale, perceived loudness, and effects of sidetone. *J*. *acoust*. *Soc*. *Amer*., 1961, 33, 160-167.

Lehiste, I. An acoustic-phonetic study of internal open juncture. *Phonetica*, Suppl. Vol. 5, 1960.

Lehiste, I. Some acoustic correlates of accent in Serbo-Croatian. *Phonetica*, 1961, 7, 114-147.

Lewis, E. N. *Experimentation on the development of more effective methods of teaching foreign languages by making extensive use of electro-mechanical aids*. Unpublished report, Contr. SAE-8715, Language Development Section, U. S. Office of Education, Washington, D. C., 1961.

Lieberman, P. Some acoustic correlates of word stress in English. *J. acoust. Soc. Amer.*, 1960, 32, 451-454.

Lindblom, B. Spectrographic study of vowel reduction. *J. acoust. Soc. Amer.*, 1963, 36, 1773-1781.

Magdics, K. Research on intonation during the past 10 years. *Acta Linguistica*, 1963, 13, 133-165.

Malmberg, B. Observations on Swedish word accent. *Stud. Ling.*, 1956, 10, 1-44.

Reichard, J. R. *Experimentation on the development of more effective methods of teaching foreign languages by making extensive use of electro-mechanical aids*. Unpublished report, Project #69, Language Development Section, U. S. Office of Education, Washington, D. C., 1962.

Rigault, A. Role de la frequence, de l'intensite et de la duree vocalique dans la perception de l'accent en francais. In: Sovijarvi, A. and Aalto, P. (eds.) *Proc. 4th Int. Congr. Phon. Sci.*, The Hague: Mouton, 1962. pp. 735-748.

Scherer. G. A. *A psycholinguistic experiment in foreign language teaching*. New York: McGraw-Hill, 1964.

Figure Captions

Fig. 1. Characteristic intonation contours of declarative sentences spoken in American English, German, Spanish and French. (From Delattre, 1962a)

Fig. 2. Experimenter's console, SAID system. The photograph shows the computer base of the system with teleprinter output (left), tape-reader input, monitoring equipment, parameter extractors, and tape recorder (extreme right).

Fig. 3. Analog electronic equipment, SAID system. (Top to bottom, left to right: test oscilloscope, mixer, frequency-to-voltage converter, monitor recorder, playback amplifier, amplifiers for parameter extraction, intensity extractor, error monitor meters, computer input amplifier, voice-operated relay, power supply, switching circuitry and filter bank for extraction of fundamental frequency, switching circuitry for mode selection, tape recorder.)

Fig. 4. Student booth SAID system. The photograph shows the loudspeakers for presentation of the pattern sentence, the microphone that receives the student's imitation, and the student console (left to right: respond light, error display, mode indicator).

Fig. 5. The difference, with respect to changes in pitch, between a triad of vowels, presented aurally, and its imitation on repeated trials by one student.

Fig. 6. The difference, with respect to changes in loudness, between the triad of vowels and its imitation. The trials are the same as those indicated in Fig. 5.

Fig. 7. The difference, with respect to changes in tempo, between the triad of vowels and its imitation. The trials are the same as those indicated in Figs. 5 and 6.

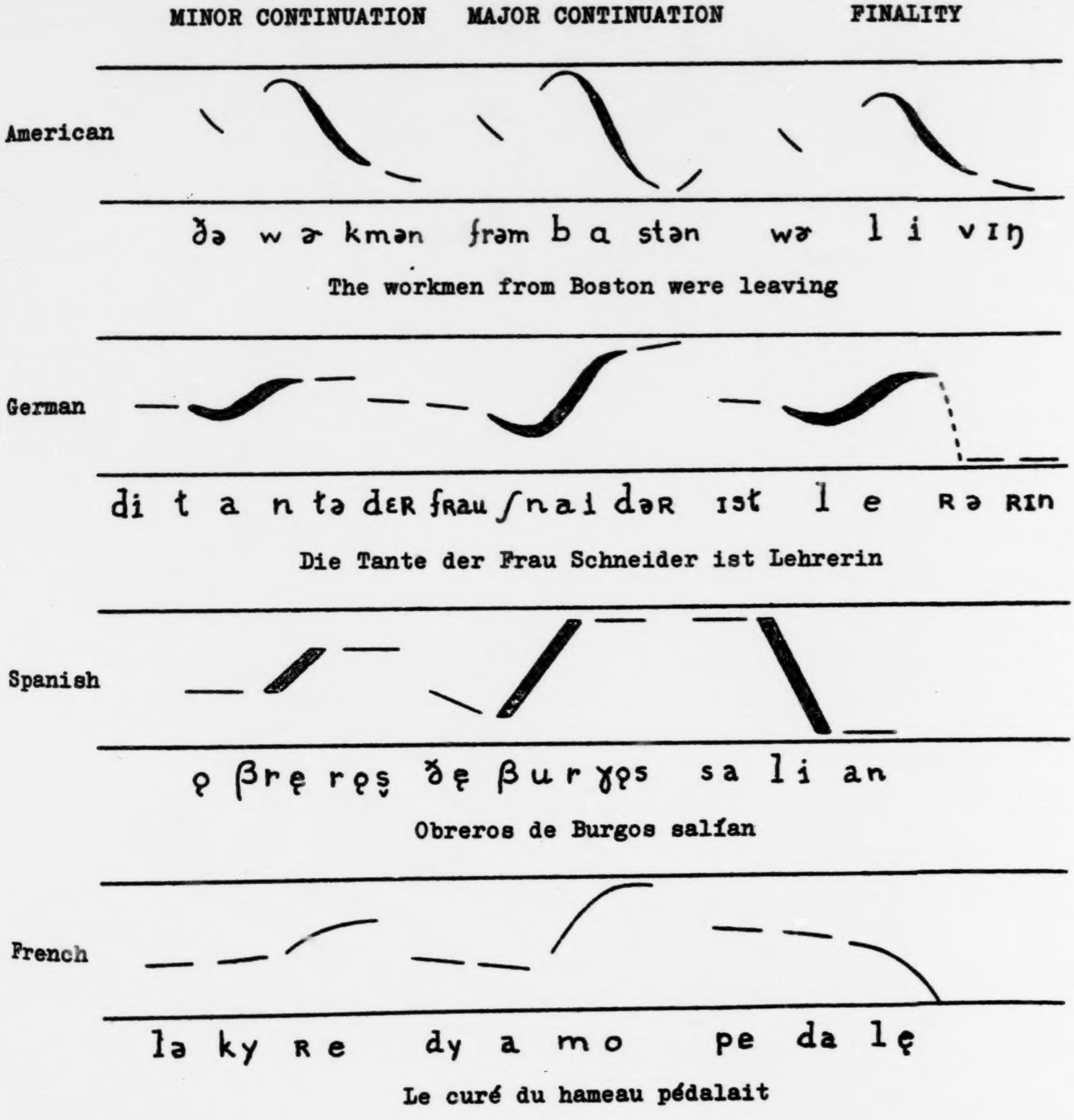

MINOR CONTINUATION
MAJOR CONTINUATION
FINALITY
American
ðə w ɚ kmən frəm b ɑ stən wɚ l i vɪŋ
The workmen from Boston were leaving
German
di t a n tə dɛʀ fʀau ʃnai dəʀ ɪst l e ʀ ə ʀɪn
Die Tante der Frau Schneider ist Lehrerin
Spanish
o̜ βre̜ ro̜s̬ ðe̜ βurɣo̜s sa li an
Obreros de Burgos salían
French
lə ky ʀe dy a mo pe da le̜
Le curé du hameau pédalait

Fig. 1

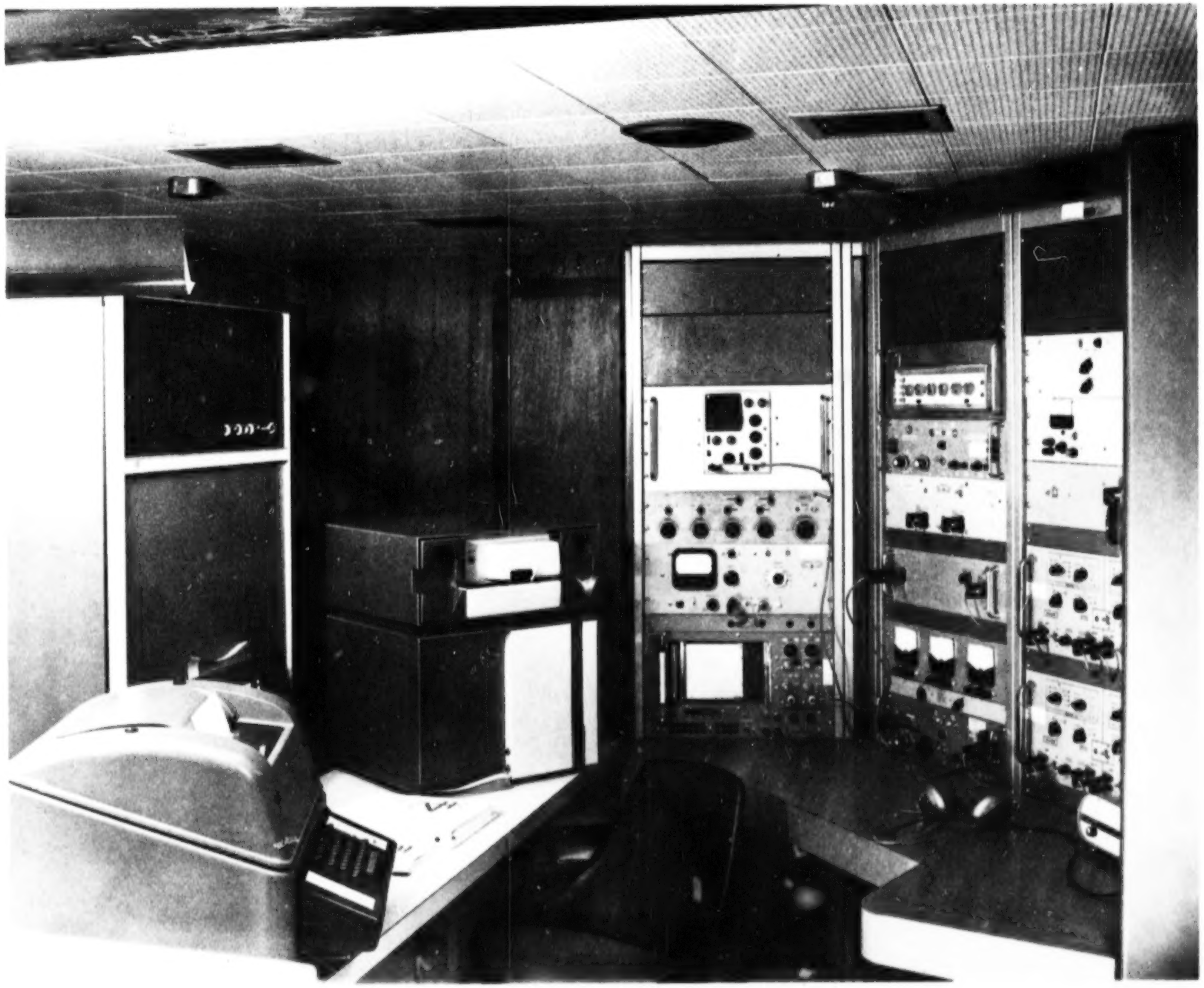


Fig. 2

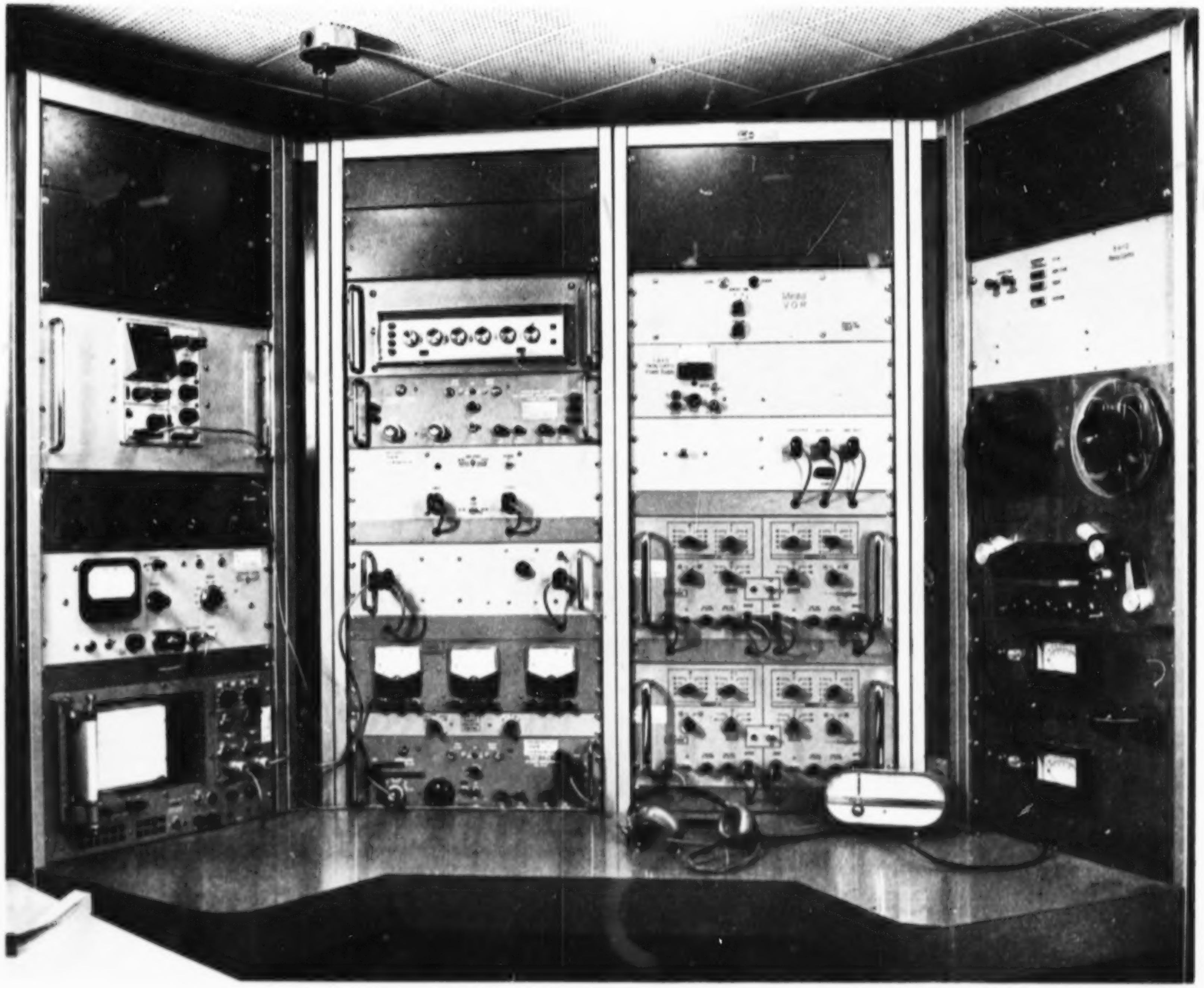


Fig. 3

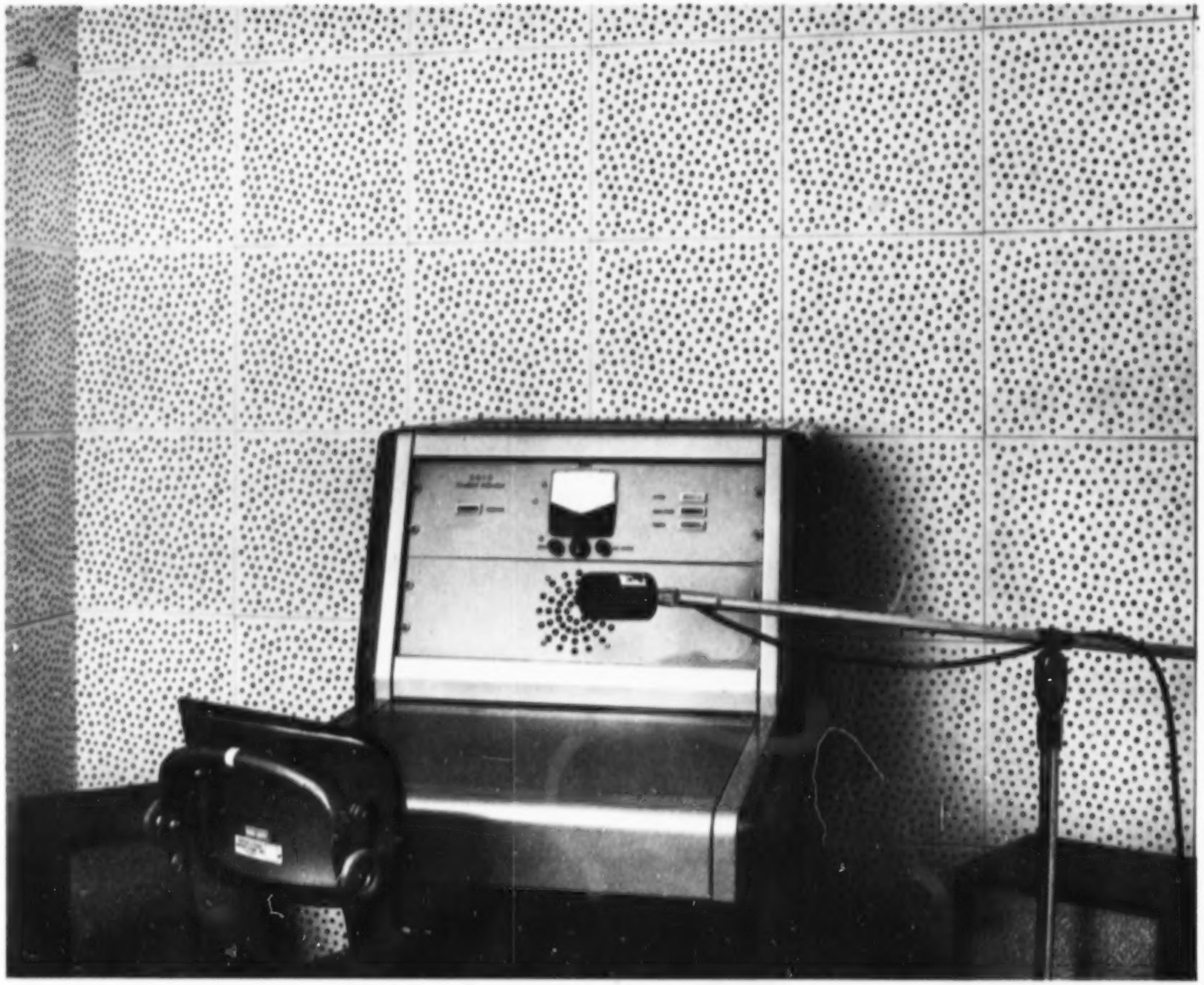


Fig. 4

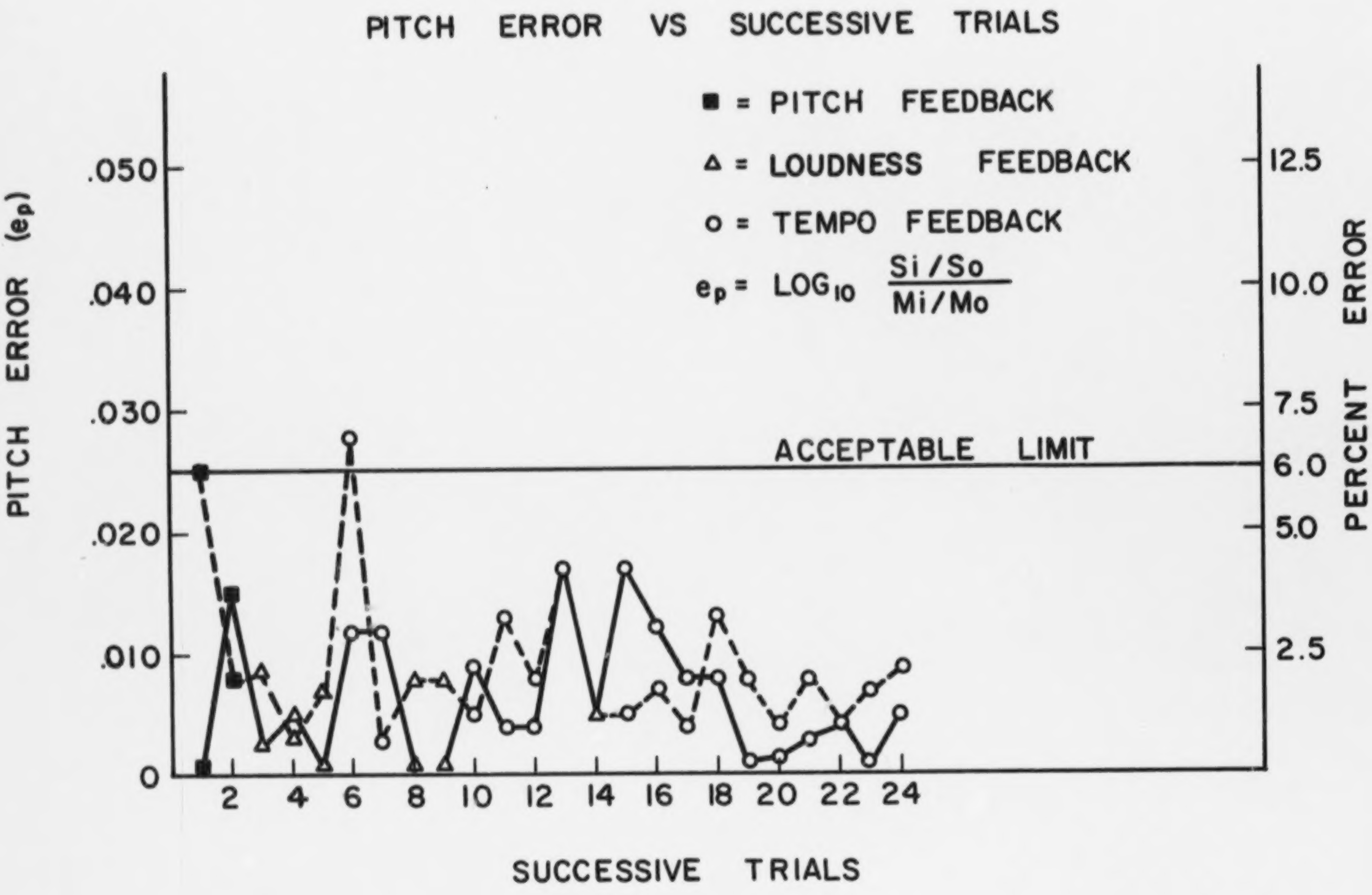

PITCH ERROR VS SUCCESSIVE TRIALS
■ = PITCH FEEDBACK
△ = LOUDNESS FEEDBACK
○ = TEMPO FEEDBACK
$e_p = \mathrm{LOG}_{10} \frac{Si/So}{Mi/Mo}$
ACCEPTABLE LIMIT
PITCH ERROR (e_p)
.050
.040
.030
.020
.010
0
2
4
6
8
10
12
14
16
18
20
22
24
SUCCESSIVE TRIALS
PERCENT ERROR
12.5
10.0
7.5
6.0
5.0
2.5

Fig. 5

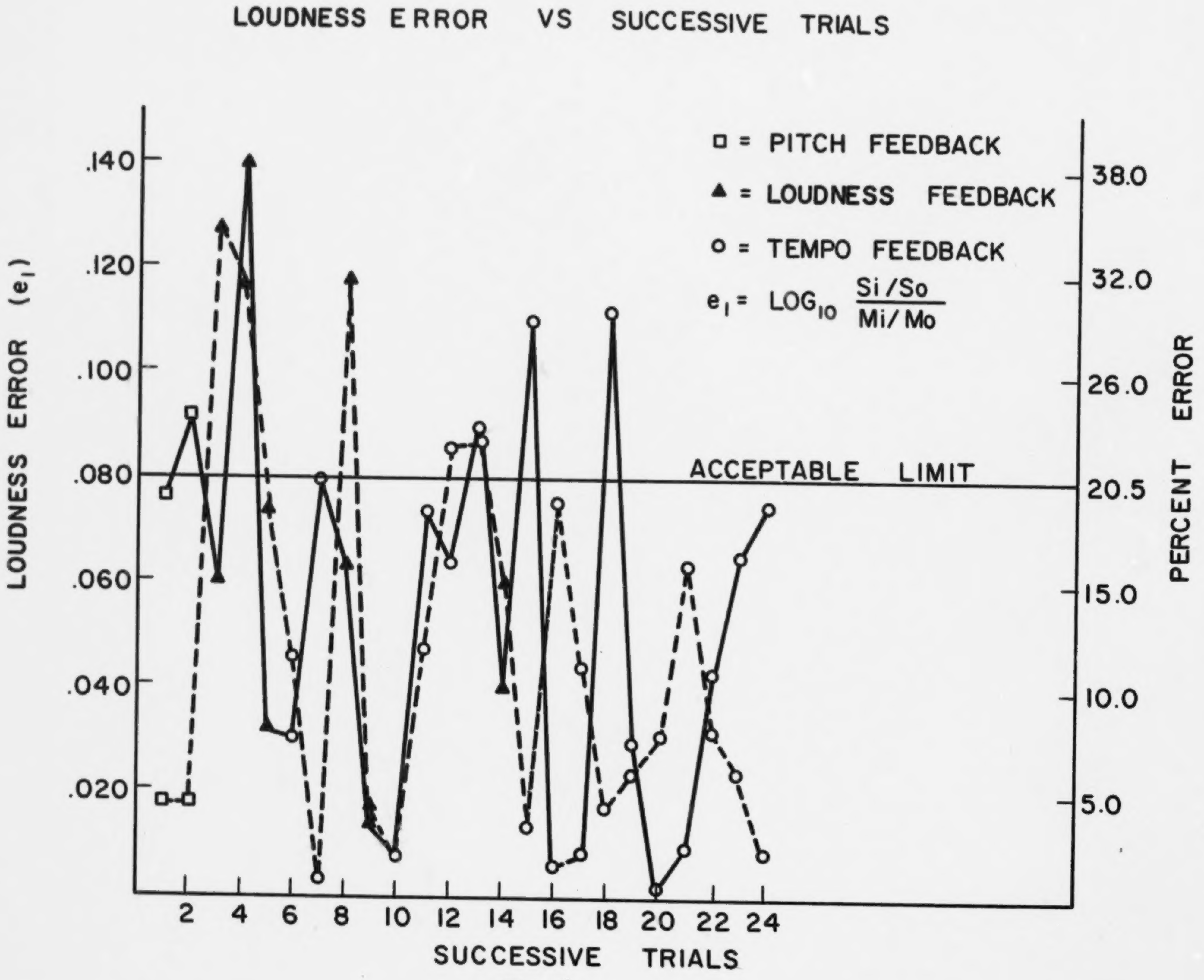

LOUDNESS ERROR VS SUCCESSIVE TRIALS
□ = PITCH FEEDBACK
▲ = LOUDNESS FEEDBACK
○ = TEMPO FEEDBACK
$e_1 = \mathrm{LOG}_{10} \frac{Si/So}{Mi/Mo}$
ACCEPTABLE LIMIT
LOUDNESS ERROR (e_1)
.140
.120
.100
.080
.060
.040
.020
PERCENT ERROR
38.0
32.0
26.0
20.5
15.0
10.0
5.0
2
4
6
8
10
12
14
16
18
20
22
24
SUCCESSIVE TRIALS

Fig. 6

TEMPO ERROR VS SUCCESSIVE TRIALS

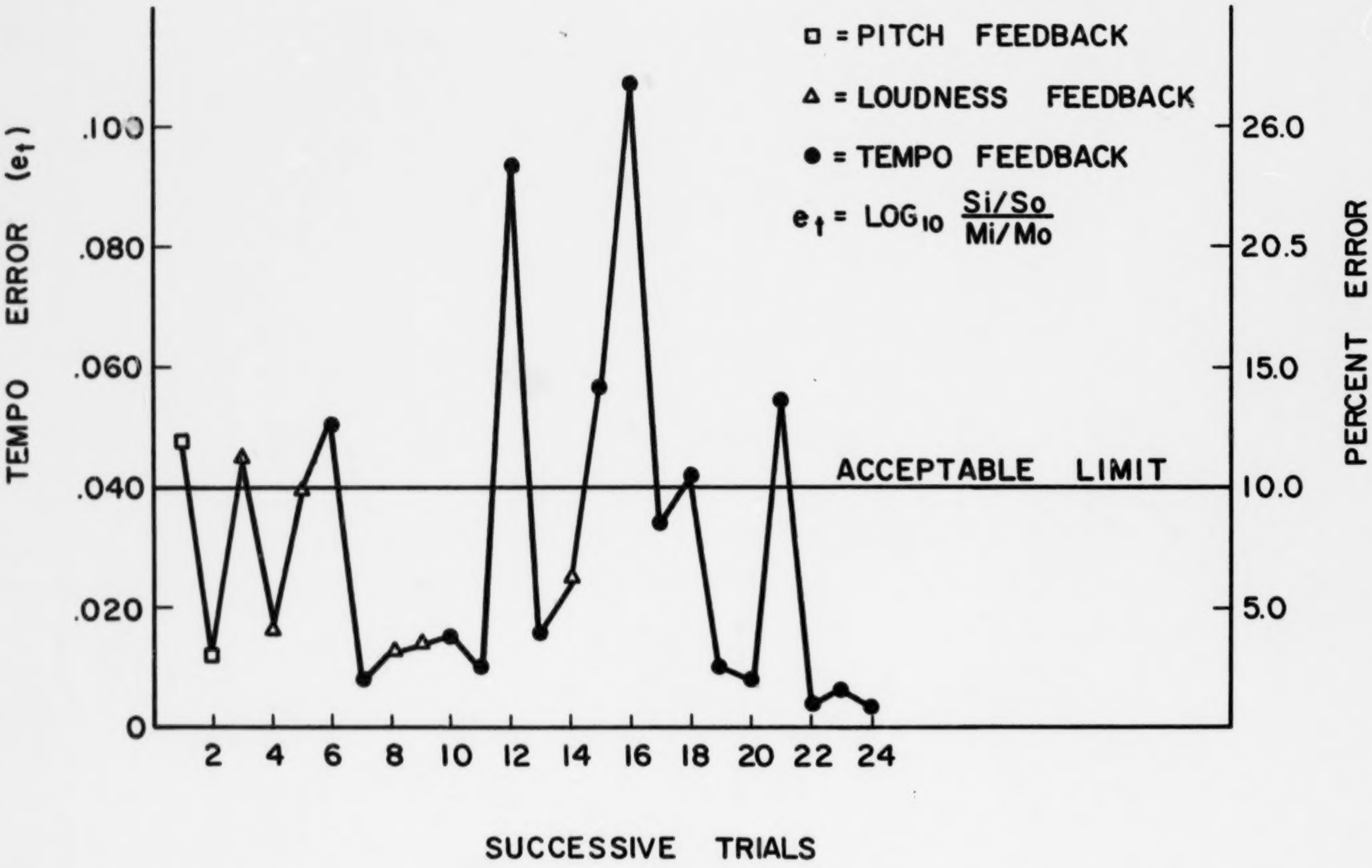


Fig. 7

A PROGRAM FOR REVIEWS AND A REVIEW OF A PROGRAM ON LINGUISTICS

Harlan Lane and George Geis
The University of Michigan

A program for reviews and a review of a program on linguistics

Harlan Lane and George Geis

A program is more like a product than it is like a book, and its review must be more like an article in Consumer's Report than like one in the Times Literary Supplement. Programs have claimed special attention in education because they are the offspring of a technology of behavior; they are generated by an iterative process of analysis--design--evaluation--revision; they embody, therefore, the principles of scientific inquiry applied to the construction of educational materials. Because the format of programs often resembles that of books, there is a danger that reviews of programs will assume the highly subjective and unscientific character of most book reviews. This outcome is inappropriate, undesirable, and unnecessary. The potential buyer of a program needs to know the hard-fact answers to certain practical questions. Stated summarily they are: What is the program supposed to do, for whom under what conditions? Does it work? What do I gain and lose in buying it?

The task of the reviewer, from our point of view, is to specify these questions in detail, as they pertain to the particular product being evaluated--and then to answer them.

It will become clear that the reviewer cannot accomplish his task without an owner's manual provided by the manufacturer of the program. If a program is not accompanied by the information in an owner's manual (the standard accompaniment of most products, from cars to ointments) this augurs ill for its proper evaluation and use, and to this degree the product is unsatisfactory. In this respect and others, programs are like tests. Both are designed to do a job, both are developed by iterative revision and evaluation, and both must be accompanied by data concerning their objectives, validity, reliability and use; otherwise, their review is infeasible, their purchase unwise.

The following review presents the questions that we believe a reviewer should ask about a program. Several lists of criteria for evaluating programs are available[1] but this article spells out, for the first time, the criteria that are peculiarly the concern of the reviewer, and it illustrates their application to a particular program: A Programmed Introduction to Linguistics (PIL)*. If this program seems pallid in the harsh light of systematic and explicit criteria, it should be noted at the outset that few other programs would do better under this scrutiny and most texts would do worse. Let that observation comfort whom it may.

Objectives

1. Are the objectives of the program stated?

The objectives can be stated with varying degrees of precision. These statements may range from brief and vague generalities, rather like those found in a college catalog, to explicit and quantitative descriptions of behavior, including terminal (or "mastery") frames in the program, or the questions and answers of a post-test.

PIL: According to a leaflet accompanying the program "the text acquaints the student with the fundamental principles of linguistic science, phonetics and phonemics." In view of the content of the program and, especially, the eleven interpolated tests, we infer that the major objectives of the program are to produce a student who can accurately complete the following tasks:

(a) Given a schematic of the vocal tract, identify the major cavities, the articulators, and the points of articulation.

---

* Cynthia D. Buchanan, A Programmed Introduction to Linguistics. Boston: D. C. Heath and Company, 1963. 288 pp. $ 4.40.

(b) State the definition of a phoneme. Given two phones in English, state whether they are in free variation or in complementary distribution.

(c) Given an articulatory description of an English phoneme (in terms of tongue height and position for the vowels, or point and manner of articulation for the consonants) write the phonemic symbol. Given the symbol, state the articulatory description. Given a simple English word, write the string of phonemic symbols that represents it.

2. Do the objectives reflect the current state of the field?

The program may undertake to teach concepts that are up-to-date, to provide the facts and skills that are required of today's (or tomorrow's) student in the particular discipline, or it may be based on an obsolete textbook or an outdated perspective of the subject matter.

PIL: The objectives are not abreast of recent developments in linguistics. The classical 'botanical' approach, followed by the program, has undergone severe criticism and revision during the past decade. To the student and to the teacher, the overall purpose of the program will surely be unclear. The program neither adopts the objectives of the phonetician, since it fails to teach the basic skills of reading and writing phonetic notation, nor the objectives of the linguist, since it fails to approach phonemics as part of the overall problem of the structure of language, that is, in the light of the interdependence among the components of a grammar.

3. Are the emphases appropriate?

The program may extensively develop concepts of central importance, relegate lesser concepts to limited treatment and omit still others whose mastery requires a disproportionate amount of time and space, or it may dwell unduly on the curios of the subject matter, the concepts that are easily mastered, or those that have short-term appeal but little long-term utility for the student.

PIL: Although the distribution of emphases cannot be judged solely by the allocation of frames to objectives, it is instructive to note that approximately 26 percent of the frames are devoted to the first objective, concerned with articulatory phonetics, 12 percent to the second objective, concerning phonemic principles, and 46 percent to the third objective, concerning the segmental phonemes of English. The remaining 16 percent is devoted about equally to introductory material and to an explanation of the supra-segmental phonemes of English. Given the scope and approach of the program, this distribution of emphases seems reasonably standard. The widely used introduction to linguistics by Gleason, for example, allocates pages to the five topics listed above in about the same proportions.

4. How do the stated objectives of the program compare with objectives typically set forth for the target population?

The program may undertake to provide all of the concepts and skills demanded of the particular population for which it is designed or it may incorporate too few of the relevant objectives to be useful to its chosen population. For example, a program that teaches sight translation from a foreign language, and that has been developed with and designed for college students, meets only a small part of the objectives normally set by teachers for this target population. If the program is used for this population at all, it will probably be supplemented extensively by other materials and techniques.

PIL: The scope of the program's objectives is distinctly limited; in this respect, the title may be misleading. An introductory course in linguistics in colleges in the United States will normally include, in addition to the major objectives of the program, some treatment of morphology and syntax, acoustic phonetics, and language history and distribution. Those who teach such courses from a traditional point of view, may find it desirable to use the program in that part of the course devoted to articulatory phonetics and English phonemics, and then to supplement the program with a textbook.

Populations

5. Is the target population described?

Just as the learner's objectives and the program's objectives must match rather closely, so, too, the learner must be sure that he belongs to the population for which the program was designed. A description of the target population of the program, like that of its objectives, may be dangerously general (e.g., "adults of all ages") or helpfully detailed. The description may range from a statement of grade level or courses that are prerequisite (again in the manner of a college catalog) to, more usefully, the identification of content skills and propaedeutic skills (the repertoire that is needed to complete this kind of program). Ideally, the target population may be delimited by students' scores on a screening test, which accompanies the program. This test may be the same as the pre-test, discussed below, which is used to evaluate the learning effected by the program, or the two tests may be different.

PIL: According to a leaflet accompanying the program, the target population includes students of speech, linguistics and language pedagogy. "No previous knowledge of linguistics is necessary to take the course." This is, of course, inadequate as a statement of the skills necessary to complete the program successfully. Students of English as a second language will find the program particularly difficult, and English speakers with dialects markedly different from that on the Eastern Seaboard of the United States will also be at a disadvantage (the inevitable result of trying to teach phonetics by means of orthography rather than by tape-recordings).

6. Is the target population appropriate for the program?

The program may demand factual knowledge and ancillary skills that its target population in fact possesses, or the inappropriateness of its requirements may be masked by their vague description. Conversely, the target population may need to know what the program offers to teach, or the program may be designed for a population that has little demand for the learning it provides.

PIL: With the reservations concerning the program's objectives stated after question 2, the target population may be considered appropriate.

7. Is the test population adequately described? Is it comparable to the target population?

The evidence that the program is effective as an educational product (see below) may not be relevant unless the population on which it was tested and the population which will use it are comparable. This cannot be decided, in turn, unless the test population, like the target population, is adequately described. Once again, performance on a reported screening test provides the most explicit statement of the entry behaviors of the students. A detailed description of content skills and propaedeutic skills is next to be preferred. The mere statement of the students' grade level, I.Q., or prior courses does not provide sufficient information to ensure that the performance of the test population is a predictor of the performance of the target population.

PIL: There is no description of a test population beyond the statement, in the foreword to the program, that undergraduate and post-graduate students completed the program in various educational contexts; the errors and comments of these students led to repeated revisions of the original program.

Conditions

8. Is there a description of the conditions under which the program should be used?

The effective use of the program may require certain environmental conditions, supplementary devices or materials, or procedures by teacher and student. The program may ensure that these conditions are met by instructions to the student on the use of the program and supplementary materials and by instructions to the teacher concerning administration and integration with classroom practices and the curriculum.

PIL: There is no teacher's manual, no statement of the conditions under which the program should be used, nor any recommendation concerning supplementary materials or devices. In a brief procedural note to the student, he is instructed to cover the answers to the frames, located in the right margin of each page, to respond to the frames orally or occasionally to write his answer, and then to check his answer by removing the cover sheet.

9. Are the conditions described under which the program was tested? Are they comparable to those conditions under which it will be used?

Unless the test and target populations used the program under comparable conditions, the effectiveness of the program with the former group is no guarantee of its success with the latter. The incentives, instructions and procedures employed with the test population, like those for the target population, must be described explicitly, or it may be unsafe to predict from the performance of one to that of the other.

PIL: No information is provided concerning the incentives, instructions and procedures employed with the test populations but these must surely have varied with such diverse groups. Since the conditions for the use of the program by the target population are also not specified, it seems likely that the two sets of conditions may be quite disparate, with a concomitant loss in the effectiveness of the program. For example, a student working on the program in a classroom, under time constraints and without the incentive to state or record his answers, may easily slip into a pattern of responding where his answer to each frame comes after, rather than before, reading the correct answer for that frame.

10. How adaptable is the program to differences in performance within the program among the individuals in the target population?

Recognizing that a clearly-defined target population is still not a homogeneous group, the program may be adaptive to differences in individual performance by permitting self-pacing, by 'branching', by including remedial subprograms, and in other ways.

PIL: The program itself makes no special provision, beyond those inherent in self-instruction (e.g., self-pacing), for individual differences in performance. However, a teacher could use the students' error rate and scores on the tests to detect deviant performance and he could then attempt to deal with these individual differences himself.

11. Are there procedures or materials for ensuring and evaluating the maintenance of the terminal behaviors?

The program may take steps to ensure the maintenance of the conditioned behavior or it may abdicate responsibility as soon as the last blank is filled in. Steps may be taken within the program itself to enhance maintenance, such as introducing sources of interference or distributing conceptually related items throughout the program; or the program may be accompanied by materials specially designed for later reconditioning and evaluation; or instructions to the teacher and student may be provided on the conduct of exercises for maintenance.

PIL: Concepts introduced early in the program recur in later sections and there is, to some measure, a hierarchical structure: responses to later frames require correct responses that have occurred long before. We may therefore speculate that there are some features of the program that contribute to maintenance. There are no auxiliary materials or instructions concerned with the maintenance of the behaviors established by the program.

12. Can the program be reused by the same student? Can it be used as a reference work or for review? Is it indexed?

The program may be designed so that the student who has completed it can work through parts of it again at a later date, or so that he can review the material before an examination, or so that he can use the program as a source of facts. On the other hand, these functions may not come within the purview of the particular program.

PIL: Provided he did not write in the answers, a student could complete the program a second time. A detailed table of contents permits quick access to particular facts within the program, facilitating review.

Stimuli and responses

13. Are the frames and answers in the program factually correct and consistent?

PIL: In many details the phonemic system for English taught by the program parallels that in the classic monograph by Trager and Smith[2]; the incisive criticisms of this formulation offered by Sledd[3] are therefore germane. The program errs occasionally in matters of fact, where traditional viewpoints have been superseded by recent research. For example, the program equivocates on the nature of stress: "Stress refers to the relative force of articulation of a given syllable" (p. 179). "Stressed syllables are louder or more prominent than unstressed syllables" (p. 179). "Unlike English, these languages have either tone or quantity accentual systems. English has a stress accentual system . . . which serves to distinguish one utterance from another by the use of stress, or relative loudness" (p.180).

The profusion of terms: stress, accent, loudness, force, prominence, bespeaks a more basic confusion of fact. Acoustic-phonetic research[4] has shown that intonation is primarily responsible for the perception of stress.

Finally, the program's notation is at times unfortunate. Consider the plight of the student who learns that the same sound "is represented phonetically by [t͡š] and phonemically by /č/" (unlike [t]--/t/, [s]--/s/, etc.) only to later find that it is also represented, perhaps most often, as recommended by the I.P.A.: [tʃ].

14. Are the kinds of cueing and their frequency of use in the program tabulated? Is frequent use made of frames for which the response is overcued?

If the program employs a variety of means, both formal and thematic, for strengthening the appropriate discriminations, and if the emission of the response depends on the learning that has gone before, the response is more likely to be available on other occasions when it is required. Some description of these properties of the program, essentially its mode of operation, would be a useful accompaniment to the programmed materials themselves. Such an analysis should involve primarily the frames' functional rather than formal characteristics. A variety of classificatory systems have been advanced, each attempting to provide descriptive categories for frames.[5] It may be excessive, however, to demand a detailed classification of frames from the author or publisher.

PIL: The kinds of cueing and their frequency of use in the program are not described. Several samples of frames taken from the beginning, middle and end of the program indicate that the student's response is frequently "overcued" in the early part of the program. Thus, the appropriate response is often a repetition of a phrase in the same or a recent frame; or it is a repetition of the response that has been confirmed in several preceding frames and is unlikely, therefore, to be more than trivially under the discriminative control of the frame. As the program progresses, there is less formal and more thematic cueing of the correct response and an increasing dependence on the learning that took place in earlier frames.

15. Are the kinds of responses and their frequency of use tabulated? Does the program call for a wide variety of responses, representative of several response classes?

The transfer of newly acquired concepts and skills is more likely to be enhanced and the scope of the students' terminal repertoire is likely to be greater if diverse appropriate responses have been reinforced under a variety of relevant stimulus conditions. Stereotypy of responding may also signal a program which is primarily teaching paired-associates rather than concepts. Of course, neither a variety of "cueing methods" nor a variety of responses insures that the program is effective.

PIL: The kinds of responses and their frequency of use are not described. Examination of the program reveals that a variety of responses and response modes are employed. Frames requiring sentence completion, labelling diagrams, filling in tables, matching terms in columns, selecting a response alternative and phonemic notation are all common. Technical descriptive terms, introduced in the program, comprise nearly half of the answers. The next largest response class consists of phonetic and phonemic symbols.

16. For most frames, are the response and the stimuli that evoke it relevant to the objectives of the frame?

Each frame may represent a discriminable step toward the overall objectives of the program (for example, toward the emission of certain correct answers on the post-test), or it may be that there are numerous redundant and irrelevant frames, or frames in which the correct response is evoked by adventitious properties of the program, such as the repeated correctness of that response in prior frames.

PIL: Typically, the frames and their answers are relevant to the overall objectives of the program. Many frames are concerned, however, with peripheral objectives that are insufficiently developed in themselves to be considered major objectives. A few frames seem to make no contribution; for example: "The frequently cited example of the Eskimo language illustrates the linguistic principle that every language provides for the expression of concepts important to the culture of its __________." (Ans.: speakers)

Research findings

17. Are the findings of a field test reported?

The program may be accompanied by a report of one or more field tests, conducted under conditions like those in which the program will be used and with a test population comparable to the target population. This information provides evidence--most important--of the effectiveness of the program. It also indicates the appropriateness of the pre- and post-tests, the variance in student performance that may be anticipated, and temporal factors in completing the program. For these purposes and others, a complete set of specifications for the program will include the following results of the field tests: (a) pre-test scores; (b) errors on learning and terminal (or mastery) frames; (c) errors on test items; (d) post-test scores; (e) time to completion. Each finding is expressed in terms of the performance of individuals or, as necessary, by measures both of central tendency and dispersion. (A more elaborate set of data might indicate retention and transfer effects.)

PIL: The findings of a field test are not reported. The leaflet states that "most students should be able to complete [the program] in fourteen hours."

18. Are the scores on a pre-test uniformly low, those on a post-test uniformly high?

Programmed learning may yield a large gain score (post-test minus pre-test) for most individuals, or the distribution of gains may be heavily skewed, or only negligible gains may be demonstrable. Of course, the appropriateness and validity of the test itself should first be examined. If the measuring instrument is inaccurate, its readings are of little use (see question 19). Assuming a valid test, low gain scores due to low pre- and post-test scores may indicate a program that is ineffective in bringing about the intended behavioral changes. If the low gain score is due to high pre-test as well as high post-test scores, the program may be underestimating the entering behaviors of the population; indeed it may not be teaching anything new to that group.

PIL: No data are reported concerning pre- and post-tests.

19. Are the pre- and post-test questions and answers presented for examination? Is there evidence of their reliability? Are they comparable? Do the tests sample the same concepts that the program teaches?

If the program is accompanied by the pre- and post-tests and evidence is provided for their reliability and validity, then high gain scores may be reasonably interpreted as evidence for the program's effectiveness. Otherwise, high gain scores may be unreliable, irrelevant, or indicative of the fact that the pre-test samples one repertoire, the post-test another. Without evidence that the program leads to an appreciable increase in the students' knowledge, as measured on suitable pre- and post-tests, the potential buyer has little reason to adopt this product.

PIL: No data are reported concerning pre- and post-tests. There is no direct evidence, therefore, that this program is an effective teaching device.

Economy

20. What is the purchase price of the program? Are there additional costs?

The publisher should indicate clearly if special aids (masks, tape recorders, teaching machines, etc.) will be needed when using the program.

PIL: The accompanying leaflet states that this paperbound program, 8 1/2 x 11, 288 pages, costs $4.40. There are no auxiliary costs.

21. Are comparable programs available?

The job of the reviewer of a program may include "comparison shopping".

PIL: Two recent program in linguistics should be mentioned but neither overlaps extensively with that under review. Harms[6] has developed a program to condition phonetic transcription of American English and Roberts[7] has published a programmed introduction to English syntax.

22. Assuming the program meets its objectives in the reported time, does it thereby reduce the overall instructional costs?

The cost analysis of education or of a particular component in education is a complicated task well beyond the scope of this paper. Even if a program is expensive, takes a long time to work through, and "covers" only a portion of a course, it still may be cheap if it invariably produces a few crucial behaviors. Such a program looks uneconomical when compared with a text that "covers" more material in less time at less cost per student. The critical questions are, however: Is there evidence that the objectives intended by the method are actually met? Are those objectives worthwhile, that is, worth the expense?

PIL: Since the program will, in most courses, be supplemented by lecture, a more comprehensive textbook, and audio-visual aids, adoption of the program will not lead to an appreciable reduction in the cost of instruction in introductory linguistics.

Footnotes

1 For example: Report of the Joint Committee on Programmed Instruction and Teaching Machines. *Audio-Visual Instruction*, 1963, 84-89; Checklist for selecting programs. *N.S.P.I. Journal*, 1963, 2, 4-8; Silver, G. M. Programmed instruction for computor programming. *Computors and Automation*, 1963, 12, 12-18.

2 Trager, G. L., & Smith, H. L., Jr. *An Outline of English structure*. Wash., D.C.: Amer. Council Learn. Soc., 1956.

3 Sledd, J. Review of Trager and Smith's "An outline of English structure." *Language*, 1955, 31, 312-345.

4 See, for example, Bolinger, D. L. A theory of pitch accent in English. *Word*, 1958, 14, 109-224.

5 E.g., Eigen, Lewis D. The construction of frames of an automated teaching program. Unpublished report, Center for Programmed Instruction, New York, 1959; Margulies, S. Some general rules of frame construction. Unpublished report, Basic Systems, Inc., New York, 1963; Markle, S. M. *Good Frames and Bad*. New York: Wiley, 1964.

6 Harms, L. S. Programmed learning for phonetic transcription. *Phonetica*, 1963, 10, 42-50. *Phonetic transcription, a programmed introduction*. Chicago: Scott-Foresman, 1964.

7 Roberts, P. *English syntax*. New York: Harcourt, Brace & World, 1964.